

AD-A119 720

ILLINOIS UNIV AT URBANA DEPT OF PHYSICS

F/G 8/3

THEORETICAL STUDIES OF THE OCEANIC INTERNAL WAVE SYSTEM.(U)

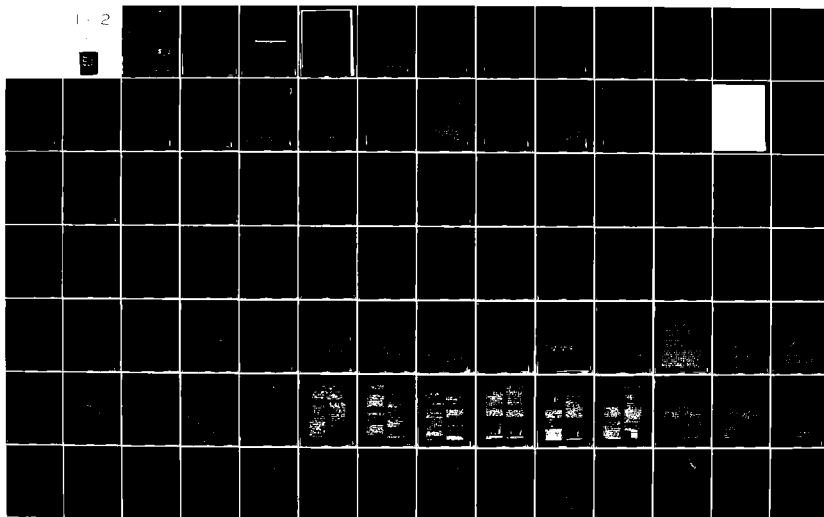
AUG 82 J WRIGHT

N00014-80-C-0840

NL

UNCLASSIFIED

1-2



AD A119720

DTIC FILE COPY

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered):

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
	AD-A119720	
4. TITLE (and Subtitle) THEORETICAL STUDIES OF THE OCEANIC INTERNAL WAVE SYSTEM		5. TYPE OF REPORT & PERIOD COVERED Final Report 06-15-80/06-30-82
7. AUTHOR(s) Jon Wright		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Illinois at Urbana-Champaign Urbana, Illinois 61801		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0840
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Branch Office Room 286, 536 South Clark St. Chicago, IL 60605		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE 08-31-82
		13. NUMBER OF PAGES
		15. SECURITY CLASS. (of this report) Unclassified
		16a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) A - approved for public release: distribution is unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)		

DD FORM 1 JAN 73 1473

82 08 30 164
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

University of Illinois at Urbana-Champaign

DEPARTMENT OF PHYSICS

Loomis Laboratory of Physics

Urbana, Illinois 61801

August 20, 1982

Dr. Lou Goodman

ONR480

Ballston Towers #1

Arlington Va. 22217

RE: N00014-80-C-840

Theoretical Studies of the Oceanic

Internal Wave System

Final report

Dear Lou:

The primary emphasis of this contract was to calculate the parameters appropriate to a quasi-linear description of the oceanic internal wave field. The main results are described in DeWitt's thesis(1). There have been two published papers (2,3) on the early part of this work. A paper describing the later calculations(4) is enclosed and will be submitted for publication shortly. We also began work on an eikonal description of high vertical wave number internal waves. That work is not yet fully finished. In addition we have been trying to understand general features of nonlinear systems with the hope that we can have a better phenomenological picture of weakly turbulent systems such as the oceanic internal wave field. That resulted in three published papers(5,6,7) plus another paper that has been submitted for publication(8).

Our calculations on the internal wave field provide better numerical values for lifetimes and other parameters than those existing calculations that relied on weak interaction theory. However the effect of the large scale internal waves on the smaller scale waves in the induced diffusion region (advection and doppler shifts) are so strong that our results are not completely reliable. In regions of wave vector space where these effects are less important but still strong enough to invalidate weak interaction theory we have results that we believe to be reliable. In particular in the region of vertical wave length from 100 to 500 meters and frequencies larger than four times the inertial frequency our decay rates are two to five times greater than weak interaction rates in spite of the claim that weak interaction theory is valid for these waves. The reason for the failure of weak interaction theory is



Approved for	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A	

that all members of a resonant triad must be long lived, not just the member whose lifetime is being calculated. In most other regions our decay rates are smaller than or equal to weak interaction rates. Our eikonal approach is an attempt to handle the region where our present results are not completely reliable.

Publications

1. DeWitt, R.J. "Self-consistent Effective Medium Parameters for Nonlinear Random Oceanic Internal Waves" Thesis University of Illinois 1982.

~~2. DeWitt, R.J. and Wright, J. "Self-consistent Methods" in Nonlinear Properties of Internal Waves 1982~~

3. DeWitt, R.J. and Wright, J. "Self-consistent Effective-medium Theory of Random Internal Waves" J. Fluid Mech. 115, 283 (1982).

4. DeWitt, R.J. and Wright, J. "Self-consistent Effective Medium Parameters for Oceanic Internal Waves" to be submitted for publication.

5. Chang, S.J., Wortis, M. and Wright, J. "Tricritical Points and Bifurcations in a Quartic Map" in Nonlinear Problems: Present and Future edited by A.R. Bishop, D.K. Campbell, and B. Nicolaenko (North-Holland, Amsterdam, 1982).

6. Chang, S.J., and Wright, J. Phys. Rev. A 23, 1419 (1981).

7. Chang, S.J., Wortis, M., and Wright, J. Phys. Rev. A 24, 2669 (1981).

8. Schult, R., and Wright, J. "Spatial Order in Chaos" submitted for publication.

Sincerely,


Jon Wright

**SELF-CONSISTENT EFFECTIVE MEDIUM PARAMETERS
FOR NONLINEAR RANDOM OCEANIC INTERNAL WAVES**

Robert Jon DeWitt, Ph.D.
Department of Physics
University of Illinois at Urbana-Champaign, 1982

In this thesis we study methods for obtaining self-consistent parameters that describe the lifetimes and frequencies of random internal waves in the ocean. We discuss previous treatments of the problem and indicate that these methods are inadequate in large part because they are not self-consistent. We perform a self-consistent calculation of oceanic parameters and compare our results to previous numbers. We end by discussing possible future improved self-consistent calculations.

Self-consistent effective-medium theory of random internal waves

By R. J. DEWITT AND JON WRIGHT

Department of Physics, University of Illinois at Urbana-Champaign,
Urbana, Illinois 61801

(Received 6 June 1980 and in revised form 11 May 1981)

In the first part of this paper we introduce a path-integral formalism for the internal-wave field of the ocean. The intent is to show that this type of formalism may be useful in suggesting improvements to current calculations, as it provides a framework for applying a wide variety of approximations that have been and are currently being developed in other areas of physics. We demonstrate the method by deriving equations for a self-consistent field approach (also known as the direct-interaction approximation). The experience in other areas of physics is that the self-consistent field approximation is more reliable than lowest-order perturbation theory. The end result of the DIA is the determination of an effective linear model for the description of internal waves in the deep-ocean environment. In the second part of the paper we obtain Hasselmann's source function by a prescribed limiting process and are able to indicate possible improvements in related calculations by comparing the limiting assumptions with numerically computed values.

1. Introduction

In the past two decades the study of the transfer of energy in the internal-wave field via three-wave (nonlinear) interactions has become an area of active research. The possibility of three-wave *resonant* interactions was first mentioned by Phillips (1977). Subsequently a Boltzmann-like equation giving energy transfer rates in the Resonant-Interaction Approximation (RIA) was obtained by Hasselmann (1968, 1967) using multiple-time-scale methods. Inherent in his calculation was the assumption that transfer times were long compared with typical oscillation periods, a reasonable assumption given the widespread belief that internal-wave interactions were weak. Later synthesis of an analytic form for the internal-wave spectrum by Garrett & Munk (1972, 1975) raised hopes that results of numerical calculations using Hasselmann's transfer equation might contain some measure of reality. Unfortunately, subsequent calculations by Olbers (1976), McComas (1977) and Pomphrey, Meiss & Watson (1980) predicted very short interaction times in a rather large region of the spectrum, so that the results were internally inconsistent. What was needed, then, was a systematic way to relax the strict assumptions inherent in the derivation of Hasselmann. The primary purpose of this paper is to describe a formalism that might very well prove useful to that end.

In recent years new methods have been developed for analysing classical random nonlinear systems which partially owe their existence to analogous methods developed earlier for quantum many-body physics and particle physics. In the classical regime

most of the emphasis has been on critical dynamics and fully developed turbulence. We propose to formulate the dynamics of both internal waves and surface waves using these methods.

We should emphasize at the outset that we make no claim that the formalism developed herein provides a complete picture of the very complicated dynamics of internal waves. Rather, our interest is in demonstrating that the formalism provides a much more general framework in which to study internal waves than do the earlier schemes. In fact the reason for mentioning surface waves at all is to demonstrate that the formalism is easily generalized from that obtained for the simple three-wave interacting system used to model internal waves. It is in part this flexibility that makes this method powerful.

The aim of our program will be to calculate the linear response functions (Green's functions) and the two-time correlation functions. The ocean will be assumed to be a stationary random system. The time-development of perturbations to the wave field is described by the linear response functions.

In the theory of turbulence there is a long history of attempts at setting up such a formalism. The early work of Kraichnan (1959) and Wyld (1961) was followed by a 1973 paper by Martin, Siggia & Rose (1973) (henceforth MSR) which introduced the idea of doubling the number of variables in the problem, leading to the current formulation. There have been a number of papers which have rewritten the MSR formalism in various ways, see for example Phythian (1977) and Langouche, Roekaerts & Tirapegui (1979). We choose to use the path-integral formalism of Phythian to set up the problem. The reason for this is that there is currently a great deal of activity in various areas of physics using path integrals. This will enable us easily to make use of some of their techniques. In this paper we will discuss perturbative and self-consistent evaluations of the path integral. In fact for the approximation we propose, we could use directly the original work of Kraichnan (1959) who proposed the direct-interaction approximation which has been extensively studied. We refer the reader to the reviews of Leslie (1973) and Orszag (1977). For a similar approximation see Phythian (1969), Herring (1965) and Edwards (1964). It is essentially a self-consistent field approximation, and, in the language of diagrams, vertex corrections are ignored. The semiclassical or WKB approximation is another standard approximation to path integrals. We will not discuss that here, but it does appear that it is the appropriate approximation for part of the internal-wave field.

In order to show how such a formalism might shed more light on previous calculations we show that a particular limiting case of our model precisely corresponds to the results of Hasselmann and consequently indicates how one might improve on the calculations of Olbers and McComas. We also show that the formalism splits Hasselmann's source function into two pieces, which have natural interpretations as effective damping and driving forces. This finally leads us to interpret our model of the internal-wave field as a collection of uncoupled, damped oscillators in the presence of random driving forces, in the spirit of effective-medium theories found in other areas of physics.

In §2 we review the derivation of the path integral for stationary random processes. In §3 we discuss the correlation functions and the linear-response functions and the direct-interaction approximation. In §4 we interpret the nonlinear problem as an effective linear theory with a memory. In §5 we discuss the relationship of our model to the results of Hasselmann. In §6 we briefly mention the WKB approximation and

the important physics that is involved. The final section is devoted to a summary and discussion of some unresolved problems.

2. Path integral

In this section we will sketch the derivation of the path-integral form of the characteristic functional describing the random field. For some of the more technical mathematical points we refer to the literature, Pythian (1977) and Langouche *et al.* (1979).

As far as is possible we will use the notation of Olbers (1976) for the mode description. The equations to be solved are of the form

$$A^s(\mathbf{k}, t) + i\Omega^s(\mathbf{k}) A^s(\mathbf{k}, t) + \nu_s(\mathbf{k}) A^s(\mathbf{k}, t) + \sum_{\mathbf{k}_1, \mathbf{k}_2} \int d^3\mathbf{k}_1 d^3\mathbf{k}_2 \delta^3(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) B_{\mathbf{k}_1 \mathbf{k}_2}^{-s} A^s(\mathbf{k}_1, t) A^s(\mathbf{k}_2, t) = f^s(\mathbf{k}, t), \quad s = \pm 1. \quad (2.1)$$

Here we have introduced the terms $\nu_s(\mathbf{k}) A^s(\mathbf{k}, t)$ and $f^s(\mathbf{k}, t)$ as external damping and driving forces on the internal-wave field. The motivation for the particular choice of these terms is twofold. The first is that of simplicity – we wish to demonstrate the application of the formalism to this problem without introducing complexities that might tend to obscure the general ideas. The second motivation can only be seen *a posteriori*. When we subsequently write down the Dyson's equations for this problem it will be seen that new terms appear which have the same formal structure as the two 'external' terms; this leads us to interpret the new terms as 'corrections' to the initial terms, i.e. as 'effective' damping and driving terms. For most modes f and ν_s are negligible and can be ignored. They are just a convenient way of parametrizing the coupling of inertial waves to the rest of the world.

The amplitudes, mode frequencies and couplings satisfy the relations:

$$A^s(\mathbf{k}, t) = [A^{-s}(-\mathbf{k}, t)]^*, \quad \Omega^s(\mathbf{k}) = \Omega^{-s}(-\mathbf{k}), \quad (2.2), (2.3)$$

$$B_{\mathbf{k}_1 \mathbf{k}_2}^{-s} = (B_{\mathbf{k}_1 - \mathbf{k}_2 - \mathbf{k}_2}^{-s})^*. \quad (2.4)$$

The index s is either $+$ or $-$ with $\Omega^+ > 0$. The label π is an additional mode index, which for internal waves is equal to unity. If an Ω appears without the label, it is taken to be Ω^+ . We make the *ad hoc* assumption that the driving forces $f_0^s(\mathbf{k}, t)$ are stationary and random with a Gaussian distribution function,

$$\langle f^s(\mathbf{k}, t) f^s(\mathbf{k}', t') \rangle = \delta^3(\mathbf{k} + \mathbf{k}') \delta_{s, -s'} R_0(\mathbf{k}, t - t'). \quad (2.5)$$

This particular assumption allows us to obtain closed-form solutions for the path integrals that will appear later. Other models for the interaction of the external forces with the internal waves could be used, such as random parametric couplings in (2.1), but such models are somewhat more difficult to treat.

The probability distributions for f are specified in terms of its characteristic functional $C(\phi)$,

$$C(\phi) = \left\langle \exp \left(i \int d^3\mathbf{k} \sum_s \int_{-\infty}^{\infty} dt \phi^{-s}(-\mathbf{k}, t) f^s(\mathbf{k}, t) \right) \right\rangle, \quad (2.6a)$$

$$C(\phi) = \exp \left\{ -\frac{1}{2} \sum_s \int d^3\mathbf{k} \int dt dt' \phi^{-s}(-\mathbf{k}, t) \phi^s(\mathbf{k}, t') R_0(\mathbf{k}, t - t') \right\}. \quad (2.6b)$$

The equations are much simpler when expressed in terms of their Fourier transforms,

$$\left. \begin{aligned} A^s(\mathbf{k}, \omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} A^s(\mathbf{k}, t) dt, \\ \phi^s(\mathbf{k}, \omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} \phi^s(\mathbf{k}, t) dt, \\ F_0(\mathbf{k}, \omega) &= \int_{-\infty}^{\infty} e^{i\omega t} R(\mathbf{k}, t) dt. \end{aligned} \right\} \quad (2.7)$$

With this convention, equation (2.6) becomes

$$C(\phi) = \exp \left\{ -\frac{2\pi}{2} \sum_s \int d^3\mathbf{k} \int d\omega \phi^{-s}(-\mathbf{k}, -\omega) F_0(\mathbf{k}, \omega) \phi^s(\mathbf{k}, \omega) \right\}. \quad (2.8)$$

We are interested in finding the probability distribution for the amplitudes A . To that end we introduce the characteristic functional for those variables,

$$Z(\xi) = \langle \exp i \sum_s \int d^3\mathbf{k} \int dt \xi^{-s}(-\mathbf{k}, t) A^s(\mathbf{k}, t) \rangle$$

or in terms of the Fourier-transformed amplitudes

$$Z(\xi) = \langle \exp [2\pi i \sum_s \int d^3\mathbf{k} \int d\omega \xi^{-s}(-\mathbf{k}, -\omega) A^s(\mathbf{k}, \omega)] \rangle. \quad (2.9)$$

The angle brackets refer to averaging over the distribution for f . They are given formally as

$$\langle X(f) \rangle = \int \mathcal{D}f W(f) X(f) \quad (2.10)$$

where $W(f)$ is the probability distribution function for f and is just the Fourier transform of C . The integral is a functional integral in that we are to integrate over all possible functions $f(\mathbf{k}, t)$ for each mode k . One way to achieve this is to discretize k and t and integrate,

$$\mathcal{D}f = \prod_k \prod_n df(k, t_n). \quad (2.11)$$

For the purposes of this paper it is not necessary to be specific and Phythian (1977) can be consulted for more details. The idea now is to replace $W(f)$ by its Fourier transform $C(\phi)$ and to change variables from f to A using (2.1):

$$\langle X \rangle = \int X \frac{\mathcal{D}f}{\mathcal{D}A} \mathcal{D}A \mathcal{D}\phi C(\phi) \exp \left\{ -i \sum_s \int d^3\mathbf{k} \int dt \phi^{-s}(-\mathbf{k}, t) f^s(\mathbf{k}, t) \right\}, \quad (2.12)$$

where it is understood that f is to be replaced by

$$\begin{aligned} f_s(\mathbf{k}, \omega) &= \{-i\omega + i\Omega(\mathbf{k}) + \nu_s(\mathbf{k})\} A^s(\mathbf{k}, \omega) \\ &+ \sum_{\mathbf{k}_1} \sum_{\mathbf{k}_2} \int d^3\mathbf{k}_1 \int d^3\mathbf{k}_2 \delta(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \delta(\omega - \omega_1 - \omega_2) \\ &\times A^s(\mathbf{k}_1, \omega_1) A^s(\mathbf{k}_2, \omega_2) B_{-12}^s. \end{aligned} \quad (2.13)$$

The driving function f has disappeared from the problem and its characteristic functional $C(\phi)$ has appeared. There are now two functional integrals, $\mathcal{D}A$ and $\mathcal{D}\phi$ and the Jacobian $\mathcal{D}f/\mathcal{D}A$ is gone. The Jacobian can be evaluated (Phythian

1977; Langouche *et al.* 1979) and is given formally by

$$\frac{\mathcal{D}f}{\mathcal{D}A} = \exp \left\{ + \int_{-\infty}^{\infty} d\omega \sum_{s, s_1} \int d^3k B_{k-s}^{s_1} A^s(0, 0) \right\}. \quad (2.14)$$

It has been shown (Langouche *et al.* 1979; Dominici & Peliti 1978) that, if diagrammatic perturbation theory is used, the Jacobian cancels 'tadpole' diagrams and both the tadpoles and the Jacobian can be omitted. See the appendix for a discussion of this point.

We now extend the definition of the angle brackets to include ϕ functionals as well as $A^s(\mathbf{k}, \omega)$ functionals,

$$\begin{aligned} \langle X(A, \phi) \rangle &= \int \mathcal{D}A \mathcal{D}\phi \frac{\mathcal{D}f}{\mathcal{D}A} X(A, \phi) C(\phi) \\ &\times \exp \left\{ -2\pi i \sum_s \int d^3k d\omega \phi^{-s}(-\mathbf{k}, -\omega) f^s(\mathbf{k}, \omega) \right\}. \end{aligned} \quad (2.15)$$

The generating functional introduced in (2.9) is now extended to include ϕ ,

$$Z(\xi, \eta) = \langle \exp \{ 2\pi i \sum_s \int d^3k \int d\omega [\xi^{-s}(-\mathbf{k}, -\omega) A^s(\mathbf{k}, \omega) + \eta^{-s}(-\mathbf{k}, -\omega) \phi^s(\mathbf{k}, \omega)] \} \rangle. \quad (2.16)$$

As we are presumably interested in correlation functions of $A^s(\mathbf{k}, \omega)$ which can be obtained by differentiation with respect to ξ , it seems unnecessary to introduce η . However, the field ϕ (and hence η) plays a crucial role in defining a response function and in simplifying the perturbation series. To demonstrate this we imagine replacing $f \rightarrow f + e$ in the equation of motion. This shows up in (2.15) as an additional factor in the integral

$$\exp \left\{ +2\pi i \sum_s \int d^3k \int d\omega \phi^{-s}(-\mathbf{k}, -\omega) e^s(\mathbf{k}, \omega) \right\}.$$

If we expand this to first order in e and consider the change in $\langle A \rangle$ we see that

$$\left\langle \frac{\delta A^s(\mathbf{k}, \omega)}{\delta e^{-s'}(-\mathbf{k}', -\omega')} \right\rangle = 2\pi i \langle A^s(\mathbf{k}, \omega) \phi^{s'}(\mathbf{k}', \omega') \rangle, \quad (2.17)$$

which can also be written

$$\left\langle \frac{\delta A^s(\mathbf{k}, t)}{\delta e^{-s'}(-\mathbf{k}', t')} \right\rangle = i \langle A^s(\mathbf{k}, t) \phi^{s'}(\mathbf{k}', t') \rangle. \quad (2.18)$$

Thus the term on the right-hand side of (2.18) gives the linear response of $A^s(\mathbf{k}, t)$ to a known perturbation. It is linear only in the sense that e is treated to first order. The rest of the dynamics is nonlinear. Response and correlation functions are calculated by taking the appropriate functional derivatives of Z , for instance

$$\langle A^s(\mathbf{k}, \omega) \phi^{s'}(\mathbf{k}', \omega') \rangle = \frac{1}{Z} \left(\frac{1}{2\pi i} \right)^2 \frac{\delta}{\delta \xi^{-s}(-\mathbf{k}, -\omega)} \frac{\delta}{\delta \eta^{-s'}(-\mathbf{k}', -\omega')} Z(\xi, \eta) \Big|_{\xi=\eta=0}. \quad (2.19)$$

The standard trick used to develop perturbation theory is to notice that, if the theory were linear, all of the integrals in (2.16) could be obtained exactly. In the actual case the factor preventing exact analytical integration is given by

$$\begin{aligned} \exp \left\{ -2\pi i \sum_{s, s_1, s_2} \int d^3k_1 d^3k_2 d^3k_3 \delta(\omega - \omega_1 - \omega_2) \delta^3(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \right. \\ \left. \times B_{k_1-k_2}^{s_2} A^s(\mathbf{k}_1, \omega_1) A^s(\mathbf{k}_2, \omega_2) \phi^{-s}(-\mathbf{k}, -\omega) \right\}. \end{aligned}$$

If we recognize that $A^s(\mathbf{k}, \omega)$ can be replaced by

$$A^s(\mathbf{k}, \omega) \rightarrow \frac{1}{2\pi i} \frac{\delta}{\delta \xi^{-s}(-\mathbf{k}, -\omega)}$$

and each ϕ by

$$\phi^s(\mathbf{k}, \omega) \rightarrow \frac{1}{2\pi i} \frac{\delta}{\delta \eta^{-s}(-\mathbf{k}, -\omega)},$$

then

$$Z(\xi, \eta) = \exp \left\{ \left(\frac{1}{2\pi} \right)^2 \int \delta(\omega_1 + \omega_2 + \omega_3) \delta^3(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) d\omega_1 d\omega_2 d\omega_3 d^3\mathbf{k}_1 d^3\mathbf{k}_2 d^3\mathbf{k}_3 \right. \\ \left. \times \sum_{s_1, s_2, s_3} B_{k_1 k_2 k_3}^{s_1 s_2 s_3} \frac{\delta}{\delta \eta^{-s_1}(-\mathbf{k}_1, -\omega_1)} \frac{\delta}{\delta \xi^{-s_2}(-\mathbf{k}_2, -\omega_2)} \frac{\delta}{\delta \xi^{-s_3}(-\mathbf{k}_3, -\omega_3)} \right\} Z_0(\xi, \eta). \quad (2.20)$$

The term $Z_0(\xi, \eta)$ is obtained from (2.15) and (2.16) by finding all of the integrals now that the nonlinear piece has been removed. The result is

$$Z_0(\xi, \eta) = \exp \left\{ \frac{2\pi}{2} \int d\omega \int d^3\mathbf{k} \sum_s \xi^{-s}(-\mathbf{k}, -\omega) \frac{F_0(\mathbf{k}, \omega)}{D^{-s}(-\mathbf{k}, -\omega) D^s(\mathbf{k}, \omega)} \xi^s(\mathbf{k}, \omega) \right\} \\ \times \exp \left\{ 2\pi i \int d\omega \int d^3\mathbf{k} \sum_s \xi^{-s}(-\mathbf{k}, -\omega) \frac{1}{D^s(\mathbf{k}, \omega)} \eta_s(\mathbf{k}, \omega) \right\}. \quad (2.21)$$

The new functions D are

$$D^s(\mathbf{k}, \omega) = \nu_0(\mathbf{k}) + i\Omega^s(\mathbf{k}) - i\omega, \quad D^+(\mathbf{k}, \omega) = [D^-(-\mathbf{k}, -\omega)]^*. \quad (2.22)$$

We can now write down a diagram expansion for the perturbation series. We do that in the next section.

In principle quantities of interest can be calculated directly from (2.15) without resorting to the perturbation theory that follows from (2.20). Currently there is a major effort aimed at extracting non-perturbative information directly from path integrals describing a wide variety of physical phenomena. As those methods tend to be difficult to implement, we propose to first test the SCFA (DIA). There are a number of quantities of interest for the internal-wave field. However, there is relatively meagre information on higher moments so in this paper we will be primarily interested in the two-point functions,

$$\frac{1}{2\pi} U^{ss'}(\mathbf{k}, \omega) \delta^3(\mathbf{k} + \mathbf{k}') \delta(\omega + \omega') = \langle A^s(\mathbf{k}, \omega) A^{s'}(\mathbf{k}', \omega') \rangle, \quad (2.23) \\ \hat{U}^{ss'}(\mathbf{k}, t) \delta^3(\mathbf{k} + \mathbf{k}') = \langle A^s(\mathbf{k}, t) A^{s'}(\mathbf{k}', 0) \rangle,$$

and

$$\frac{1}{2\pi} G^{ss'}(\mathbf{k}, \omega) \delta^3(\mathbf{k} + \mathbf{k}') \delta(\omega + \omega') = i \langle A^s(\mathbf{k}, \omega) \phi^{s'}(\mathbf{k}', \omega') \rangle, \quad (2.24) \\ \hat{G}^{ss'}(\mathbf{k}, t) \delta^3(\mathbf{k} + \mathbf{k}') = i \langle A^s(\mathbf{k}, t) \phi^{s'}(\mathbf{k}', 0) \rangle.$$

The definitions of U and G lead to some important symmetry properties:

$$U^{ss'}(\mathbf{k}, \omega) = [U^{-s-s'}(-\mathbf{k}, -\omega)]^*, \quad U^{ss'}(\mathbf{k}, \omega) = U^{s's}(-\mathbf{k}, -\omega), \\ G^{ss'}(\mathbf{k}, \omega) = [G^{-s-s'}(-\mathbf{k}, -\omega)]^*.$$

3. Perturbation theory

In this section we will exhibit the general structure of the two-point functions and discuss briefly the diagrammatic perturbation theory. We begin by listing the free Green's functions and correlation functions, and we will then point out the general features that the exact functions will have. The free correlation function is defined by

$$\begin{aligned} \frac{1}{2\pi} U_0^s(\mathbf{k}, \omega) \delta^s(\mathbf{k} + \mathbf{k}') \delta(\omega + \omega') \delta_{s,-s'} \\ = \langle A^s(\mathbf{k}, \omega) A^{s'}(\mathbf{k}', \omega') \rangle_0 \\ = \frac{1}{Z_0} \left(\frac{1}{2\pi i} \right)^2 \frac{\delta}{\delta \xi^s(-\mathbf{k}, -\omega)} \frac{\delta}{\delta \xi^{s'}(-\mathbf{k}', -\omega')} Z_0 \Big|_{\xi=\eta=0} \\ = \frac{1}{2\pi} \frac{1}{D^s(-\mathbf{k}, -\omega)} F_0(\mathbf{k}, \omega) \frac{1}{D^{s'}(\mathbf{k}, \omega)} \delta^s(\mathbf{k} + \mathbf{k}') \delta(\omega + \omega') \delta_{s,-s'}, \quad (3.1) \\ U_0^s(\mathbf{k}, \omega) = \frac{F_0(\mathbf{k}, \omega)}{v_0^s(\mathbf{k}) + [\omega - \Omega^s(\mathbf{k})]^2}. \end{aligned}$$

The free-response function is given by

$$\begin{aligned} \frac{1}{2\pi} G_0^s(\mathbf{k}, \omega) \delta^s(\mathbf{k} + \mathbf{k}') \delta(\omega + \omega') \delta_{s,-s'} \\ = i \langle A^s(\mathbf{k}, \omega) \phi^{s'}(\mathbf{k}', \omega') \rangle \\ = i \left(\frac{1}{2\pi i} \right)^2 \frac{1}{Z_0} \frac{\delta}{\delta \xi^s(-\mathbf{k}, -\omega)} \frac{\delta}{\delta \eta^{s'}(-\mathbf{k}', -\omega')} (Z_0) \Big|_{\xi=\eta=0} \\ = \frac{1}{2\pi} \frac{1}{v_0^s(\mathbf{k}) + i\Omega^s(\mathbf{k}) - \omega} \delta^s(\mathbf{k} + \mathbf{k}') \delta(\omega + \omega') \delta_{s,-s'}. \quad (3.2) \end{aligned}$$

If F_0 is independent of ω , the corresponding time-dependent functions are

$$U_0^s(\mathbf{k}, t) = \frac{F_0(\mathbf{k})}{2v_0^s(\mathbf{k})} e^{i\omega_0 t} [\theta(t) e^{-\gamma t} + \theta(-t) e^{\gamma t}] \quad (3.3)$$

and

$$G_0^s(\mathbf{k}, t) = e^{-i\omega_0 t} e^{-\gamma |t|} \theta(t). \quad (3.4)$$

The exact functions will have the same structure as that exhibited in (3.1) and (3.2), except that F_0 , v_0 , and Ω will be replaced by new functions that depend on ω . To the extent that the ω dependence of the effective f , v and Ω can be ignored, (3.3) and (3.4) are exact. The ω dependence will modify at least the short-time behaviour. If the singularity nearest the real ω axis remains a simple pole, the long-time behaviour will have the structure of (3.3) and (3.4). We now proceed to develop the perturbation scheme to calculate the effective f , v and Ω using standard diagrammatic techniques. We list the rules for calculating diagrams:

(a) There are two kinds of propagators, one corresponding to correlation functions and the other to response functions. Each has double and single lines. The response functions have an extra arrow pointing out from the vertex and we will conventionally take the quantity ϕ as the response function. The correlation functions have no arrow pointing out from the vertex and

(b) Each vertex has two lines pointing toward the vertex and

$$\frac{i}{D^s(k, \omega)} = \frac{s, k, \omega}{\longrightarrow}$$

$$\frac{-F(k)}{D^{-s}(-k, -\omega) D^s(k, \omega)} = \frac{s, k, \omega}{\longrightarrow}$$

FIGURE 1. Lines representing response function and correlation function.

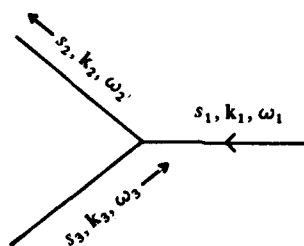
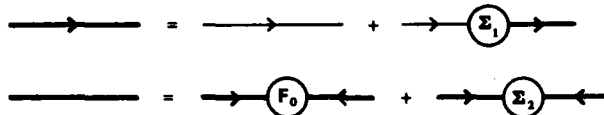
FIGURE 2. Lowest-order vertex function. The line with the incoming arrow is the first argument $B_{k_1-k_2-k_3}^{s_1-s_2-s_3} \times \delta^3(k_1-k_2+k_3) \times \delta(\omega_1-\omega_2+\omega_3)$.FIGURE 3. Lowest-order graph contributing to the correlation function, U . Because of its symmetry there is a factor of $\frac{1}{2}$ associated with it.

FIGURE 4. Dyson functions for the two-point functions.

two other lines that either have arrows pointing away from the vertex or no arrows. In figure 2 we show a vertex which has associated with it the factor

$$B_{k_1-k_2-k_3}^{s_1-s_2-s_3} \delta^3(k_1-k_2+k_3) \delta(\omega_1-\omega_2+\omega_3).$$

The first pair corresponds to the arrow in the diagram. B is symmetric in the other pairs.

(c) All internal s, k and ω 's are summed over.

(d) For N th-order perturbation theory, draw all topologically distinct diagrams. Graphs possessing a symmetry have a symmetry factor; for example the graph of figure 3 gets a factor of $\frac{1}{2}$.

(e) Each diagram gets a factor $(2\pi)^{-k(N+N_e)}$.

(f) Each external A field gets a factor of $(-i)$. (The extra i here is because the response function is $i\langle A\phi \rangle$, see (3.2).)

We mention that, if vertices with more lines are present, the only rule that is changed is (e). Each vertex will still have only one line with an arrow pointing into it.

The Dyson equations have been given by MSR. For completeness we show them in figures 4-6. Note that, although there is only one kind of bare vertex, there are three kinds of renormalized vertices which are shown to lowest order in figure 6.

$$\begin{aligned}\Sigma_1 &= \text{loop with vertex 1} + \frac{1}{2} \text{loop with vertex 2} \\ \Sigma_2 &= \text{loop with vertex 1} + \text{loop with vertex 2} + \frac{1}{2} \text{loop with vertex 3}\end{aligned}$$

FIGURE 5. Equation for Σ_1 and Σ_2 in terms of renormalized vertices.

$$\begin{aligned}\text{vertex 1} &= \text{bare vertex} + \text{triangle diagram} + \text{triangle diagram} \\ &\quad + \text{triangle diagram} + \dots \\ \text{vertex 2} &= \text{triangle diagram} + \text{triangle diagram} + \dots \\ \text{vertex 3} &= \text{triangle diagram} + \dots\end{aligned}$$

FIGURE 6. First terms in expansion of vertices.

Finally, it can easily be shown that any graph containing a closed loop whose lines are entirely response functions (arrows) is identically zero.

The Dyson equations are easily solved to give

$$G^{ss'}(\mathbf{k}, \omega) = [D^{-s}(\mathbf{k}, \omega) \delta_{s,-s'} - i \Sigma_1^{ss'}(\mathbf{k}, \omega)]^{-1} \quad (3.5)$$

(note that this is a 2×2 matrix equation). The correlation function is given by

$$U^{ss'} = G^{ss_1}(\mathbf{k}, \omega) [F_s(\mathbf{k}, \omega) \delta_{s_1,-s_2} + \Sigma_2^{ss_1}(\mathbf{k}, \omega)] G^{s_2 s'}(-\mathbf{k}, -\omega) \quad (3.6)$$

(implies summation over s_1, s_2).

If vertex corrections are ignored (direct interaction approximation) then we have

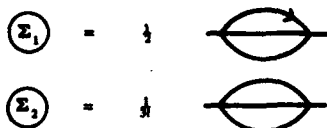
$$\begin{aligned}\Sigma_1^{ss'}(\mathbf{k}, \omega) &= \frac{2}{\pi} \sum_{s_1, s_2} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 d\omega_1 d\omega_2 \delta^s(\mathbf{k} + \mathbf{k}_2 - \mathbf{k}_1) \delta(\omega + \omega_2 - \omega_1) \\ &\quad \times B_{s_1 s_2}^{-s_1 s_2} B_{s_1' s_2'}^{-s_1' s_2'} [i G^{s_1 s_1'}(\mathbf{k}_1, \omega_1)] [-U^{s_2 s_2'}(\mathbf{k}_2, \omega_2)] \quad (3.7)\end{aligned}$$

and

$$\begin{aligned}\Sigma_2^{ss'}(\mathbf{k}, \omega) &= \frac{1}{\pi} \sum_{s_1, s_2} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 d\omega_1 d\omega_2 \delta^s(\mathbf{k} + \mathbf{k}_2 + \mathbf{k}_1) \delta(\omega + \omega_2 + \omega_1) \\ &\quad \times B_{s_1 s_2}^{-s_1 s_2} B_{s_1' s_2'}^{-s_1' s_2'} U^{s_1 s_1'}(\mathbf{k}_1, \omega_1) U^{s_2 s_2'}(\mathbf{k}_2, \omega_2); \quad (3.8)\end{aligned}$$

Σ_1 has the same symmetry properties as \mathbf{U} and $i\Sigma_1$ has the same as \mathbf{G} .

One of the assumptions of weak-interaction theory is that the diagonal elements of \mathbf{U} and \mathbf{G} can be ignored. \mathbf{U} and \mathbf{G} are assumed to have the form $U^{ss'} = U^s \delta_{s,-s'}$. The corresponding statement for \mathbf{G} is $G^s = G^{-s}$.

FIGURE 7. Graphs for Σ for four-wave interactions.

We now comment briefly on the application of these techniques to surface waves. The principal difference is that the equations involve four-wave coupling, that is there is a term

$$B \times A \times A \times A$$

in the equation of motion, (1.1). The term involving only two A 's is still there, but it is not possible to have three-wave resonances. This means that in (3.12) the limit $\nu \rightarrow 0$ gives zero and the answer must come from finite ν which presumably gives a small contribution. In any case if we replace the BAA terms by $BAAA$ in (2.1) we obtain a very similar theory. The graphs for Σ are shown in figures 7 and they are to be solved self-consistently with the diagrams of figure 4. The number associated with a vertex is now $3!$ instead of $2!$. There are some additional (2π) factors. The net effect is to change rule (c),

$$(2!)^N (2\pi)^{-\frac{1}{2}(N+N_0)} \rightarrow (3!)^N (2\pi)^{-\frac{1}{2}N_0+N},$$

and rule (b) gets a factor of i for each vertex. These two changes can be stated in a way that is easily generalized. If there are N_A A fields at a vertex, the vertex has associated with it $(i^{N_A+1} N_A!)$. Each closed loop has associated with it an integral $\int d\omega/2\pi$.

4. Linear effective medium

In this section we shall elaborate on the effective medium picture and discuss the direct-interaction approximation and the relationship to the theory due to Hasselmann. We will first remind the reader of a few simple properties of linear systems, and compare these with our effective medium. The simplest situation assumes no memory and is given by

$$A'' + i\Omega A' + \nu A' = f''(t), \quad (4.1)$$

which is (2.1) with $B=0$ and the mode index k ignored. If we define

$$\hat{U}(t, t') = \langle A^*(t) A(t') \rangle, \quad (4.2)$$

then in the steady state U depends only on $t-t'$ and is given by

$$\hat{U}(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{F e^{i\omega t}}{(\nu - i\Omega + i\omega)(\nu + i\Omega - i\omega)}. \quad (4.3)$$

If we calculate $dU(t, t)/dt$ we obtain

$$d\hat{U}(t, t)/dt = -2\nu\hat{U}(0) + F = 0, \quad (4.4)$$

where F is given by (2.5) and (2.7). This just represents the balance between the driving forces and the dissipative forces. We now compare with the correlation function obtained in the previous section,

$$\hat{O}^*(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{i\omega t} \left[\frac{1}{\nu_0 - i\Omega + i\omega - i\Sigma_1(-\omega)} \right] (F_0 + \Sigma_2) \left[\frac{1}{\nu_0 + i\Omega - i\omega - i\Sigma_1^*(\omega)} \right]. \quad (4.5)$$

We have suppressed the mode index k . If Σ_1, Σ_2 are approximately independent of ω , then it is clear that real Σ_1 is a frequency shift and $\mathcal{J}\Sigma_1$ is an effective viscosity representing loss to the other modes. This particular interpretation may be somewhat misleading, however. Consider the case where the mode in question is Doppler-shifted because of passage through an inhomogeneous medium (large-scale wave). In this case we would formally say that one mode has been damped while another mode of neighbouring wavenumber has grown. In reality of course we are just seeing the original wave with a slightly different frequency.

In addition to the 'effective damping' term there is also an effective driving force due to the other modes and its correlation function is Σ_2 . The correlation function for the total driving force is thus $\Sigma_2 + F_0$.

Now in general Σ_1 and Σ_2 will depend on ω . In this case it is easy to see that the equation giving the evolution of the amplitudes is given by

$$A(t) + \int_{-\infty}^{\infty} \hat{\Gamma}(t-t') A(t') dt' = f(t), \quad \hat{\Gamma}(t) = 0, \quad t < 0. \quad (4.6)$$

If we write

$$\Gamma^+(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} \hat{\Gamma}(t) dt, \quad \Gamma^-(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} \hat{\Gamma}^*(t) dt, \quad (4.7)$$

then we see that $\Gamma^\pm(\omega)$ is analytic in ω for $\mathcal{J}\omega > 0$. The corresponding response functions are

$$G^\pm(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d\omega e^{-i\omega t}}{\Gamma^\pm(\omega) - i\omega}, \quad (4.8)$$

which leads to an obvious interpretation of $\Sigma_1^\pm(\omega)$,

$$\Gamma^+(\omega) = \nu_0 + i\Omega - \Sigma_1^+(\omega). \quad (4.9)$$

The fact that $G(t) = 0$ for $t < 0$ (causality) implies that there are no zeros of $\Gamma(\omega) - i\omega$ in the upper half-plane. Because of these analytical properties of G and Γ , Kramer's Krönig dispersion relations can be used to calculate the real part of G and Σ once the imaginary part is known (or vice versa).

The equilibrium condition relating the energy, dissipation and driving forces is obtained by evaluating

$$U(t) = \int_{-\infty}^{+\infty} e^{-i\omega t} \frac{F(\omega)}{2\pi} \frac{d\omega}{[\Gamma^-(\omega) - i\omega][\Gamma^+(\omega) - i\omega]}$$

at $t = 0$.

5. Relationship to previous results

In this section we obtain the results of Hasselmann as a special case of our model equation. The limiting assumptions made in obtaining the results provide a more transparent way to check the validity of the assumptions inherent in Hasselmann's equation; in the end of this section we indicate where improvements could be made in his scheme.

To obtain the results of Hasselmann one performs the following operations on (3.7) and (3.8) (using only $\Sigma^0 = \Sigma^{-\infty}$).

(1) Assume that F_0 , Σ_1 , and Σ_2 are independent of ω and perform the ω_1 and ω_2 integrations. For example, defining $\nu(\mathbf{k}) \equiv \nu_0(\mathbf{k}) + \mathcal{J}\Sigma_1(\mathbf{k})$, we get

$$\begin{aligned} & \int d\omega_1 d\omega_2 \delta(\omega - \omega_1 - \omega_2) [\nu(\mathbf{k}_1)^2 + (\Omega(\mathbf{k}_1) - \omega_1)^2]^{-1} [\nu(\mathbf{k}_2) + i\Omega(\mathbf{k}_2) - i\omega_2]^{-1} \\ &= \frac{\pi i}{\nu(\mathbf{k}_1)} [\omega - \Omega(\mathbf{k}_1) - \Omega(\mathbf{k}_2) - i(\nu(\mathbf{k}_1) + \nu(\mathbf{k}_2))] \\ & \quad \times [(\omega - \Omega(\mathbf{k}_1) - \Omega(\mathbf{k}_2))^2 + (\nu(\mathbf{k}_1) + \nu(\mathbf{k}_2))^2]^{-1}. \end{aligned} \quad (5.1)$$

(2) Take the limit F_0 , ν_0 , Σ_1 , $\Sigma_2 \rightarrow 0$ (in such a way that $\hat{U}(\mathbf{k}, 0) = F/2\nu$ is maintained). Then (5.1) becomes

$$\frac{\pi^2}{\nu(\mathbf{k}_1)} \delta(\omega - \Omega(\mathbf{k}_1) - \Omega(\mathbf{k}_2)) + \frac{i\pi}{\nu(\mathbf{k}_1)} \frac{\mathcal{P}}{\omega - \Omega(\mathbf{k}_1) - \Omega(\mathbf{k}_2)}, \quad (5.2)$$

where \mathcal{P} represents a principal part integral. When the actual structure of the B coefficients is considered it is easily seen that the delta function contributes to $\nu(\mathbf{k})$ and the principal part to a frequency shift. The Dirac delta function has come from the limit

$$\frac{\nu}{\nu^2 + (\Delta\Omega)^2} \xrightarrow{\nu \rightarrow 0} \pi \delta(\Delta\Omega) \quad (5.3)$$

and we call the function on the left-hand side the frequency filter for the problem. Such a function was earlier obtained by Holloway & Henderson (1977) and Holloway (1978, 1979), where he also noted the RIA as a limiting case. This is the first indication of a possible improvement in the calculations of Olbers and McComas. What they did, in essence, was to assume that ν was small compared with typical oscillation frequencies and then calculated values of ν on the order of or even greater than the frequencies. The earlier assumption of neglecting the diagonal elements of \mathbf{U} , \mathbf{G} corresponds approximately to assuming $\nu/\Omega \ll 1$.

Equations (3.7) and (3.8) must be solved self-consistently, however, so that values of $\nu(\mathbf{k})$ used for input on the right-hand side are the same as those obtained on the left-hand side. Unfortunately solving these equations self-consistently does involve integration over one more variable, and so may require an inordinate amount of computer time (the two-dimensional integral is already relatively time-consuming).

Having now carried out the limiting process, we are ready to make a detailed comparison with the source function of Hasselmann, which he writes as

$$\begin{aligned} \mathcal{S} = & \int d^3\mathbf{k}_1 d^3\mathbf{k}_2 (T + \delta^2(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \delta(\Omega - \Omega_1 - \Omega_2) [n_1 n_2 - n n_1 + n n_2] \\ & + 2T - \delta^2(\mathbf{k} - \mathbf{k}_1 + \mathbf{k}_2) \delta(\Omega - \Omega_1 + \Omega_2) [n_1 n_2 + n n_1 - n n_2]), \end{aligned} \quad (5.4)$$

where

$$n(\mathbf{k}) \delta(\mathbf{k} + \mathbf{k}') = \langle A^+(\mathbf{k}, t) A^-(\mathbf{k}', t) \rangle \frac{1}{\Omega(\mathbf{k})}.$$

We split \mathcal{S} into two pieces,

$$\begin{aligned} \mathcal{S}_2(\mathbf{k}) = & \int d^3\mathbf{k}_1 d^3\mathbf{k}_2 (T + \delta^2(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \delta(\Omega - \Omega_1 - \Omega_2) n_1 n_2 \\ & + 2T - \delta^2(\mathbf{k} - \mathbf{k}_1 + \mathbf{k}_2) \delta(\Omega - \Omega_1 + \Omega_2) n_1 n_2) \end{aligned} \quad (5.5)$$

and

$$\mathcal{S} = \mathcal{S}_2 - \mathcal{S}_1.$$

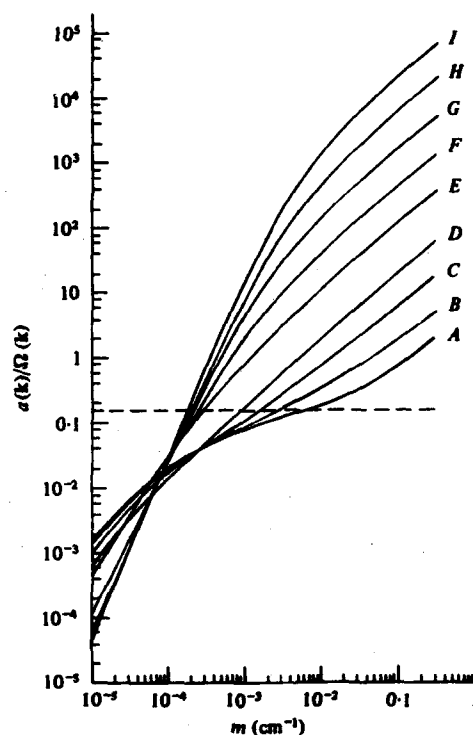


FIGURE 8. Plots of $a(k)/\Omega$ vs. m for various frequencies. The values of Ω for each curve are as follows: (A) $7.0 \times 10^{-5} \text{ s}^{-1}$, (B) $7.3 \times 10^{-5} \text{ s}^{-1}$, (C) $8.0 \times 10^{-5} \text{ s}^{-1}$, (D) $1.1 \times 10^{-4} \text{ s}^{-1}$, (E) $1.7 \times 10^{-4} \text{ s}^{-1}$, (F) $3.2 \times 10^{-4} \text{ s}^{-1}$, (G) $6.2 \times 10^{-4} \text{ s}^{-1}$, (H) $1.2 \times 10^{-3} \text{ s}^{-1}$, (I) $2.2 \times 10^{-3} \text{ s}^{-1}$. This labelling applies to all curves in figures 8 and 9. The values of Ω correspond to equal increments in $\log [(k_1^2 + k_2^2)/k_3^2]$.

We find the following identifications,

$$\frac{1}{n(k)} \mathcal{S}_2(k) = \frac{\Sigma_2(k)}{\hat{U}(k, 0)} \quad (5.6)$$

and

$$\frac{1}{n(k)} \mathcal{S}_1(k) = 2(\nu(k) - \nu_0(k)) = 2\mathcal{S}\Sigma_1(k), \quad (5.7)$$

where \hat{U} is defined in (2.23), Σ_2 is the effective force due to all other modes, and F is the total force,

$$F(k) = F_0(k) + \Sigma_2(k). \quad (5.8)$$

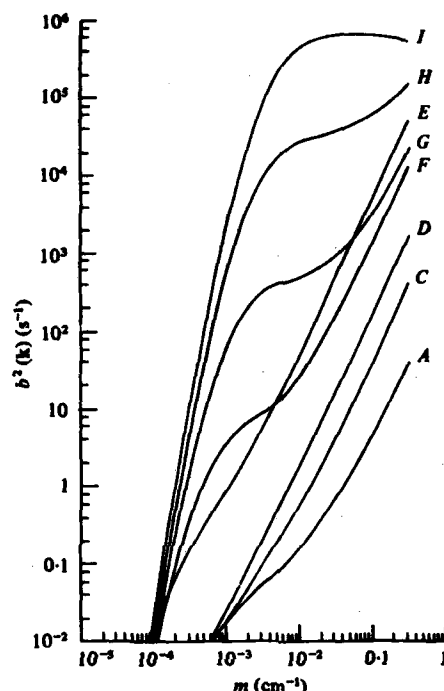
From equation (4.4) we have, in the stationary state,

$$F = +2\nu \hat{U}(k, 0) = F_0(k) + \Sigma_2(k)$$

or

$$\frac{F_0(k)}{\hat{U}(k, 0)} = 2\mathcal{S}\Sigma_1(k) - \frac{\Sigma_2(k)}{\hat{U}(k, 0)} + 2\nu_0(k), \quad (5.9)$$

$$\frac{F_0(k)}{\hat{U}(k)} - 2\nu_0(k) = \frac{-\mathcal{S}(k)}{n(k)} = \frac{\mathcal{S}_1(k)}{n(k)} - \frac{\mathcal{S}_2(k)}{n(k)}. \quad (5.10)$$

FIGURE 9. Plots of $b^2(k)$ vs. m for various frequencies.

From this equation we see that the calculations of Olbers and McComas give the coupling to external systems in terms of the difference between an effective force and an effective viscosity.

Earlier we noted the breakdown of the resonant-interaction approximation (RIA) because of the large values of $\nu(k)$ in some regions of the internal-wave spectrum. We would next like to test the assumptions that Σ_1 and Σ_2 are independent of ω . We consider first the case of Σ_1 . In order to ascertain whether the non-resonant structure of $\Sigma_1(\omega)$ might have any appreciable effect on the values computed in the right-hand side of (3.7) we choose a particularly simple form of Σ_1 :

$$-i\Sigma_1(k, \omega) = a(k) + (\omega - \Omega(k))b(k), \quad (5.11)$$

where both $a(k)$ and $b(k)$ are assumed real. If the $\Sigma_1 = \text{constant}$ assumption is warranted we should expect to compute very small values for $b(k)$. For the present we still assume that $\Sigma_2 = \text{constant}$ and use the RIA. The effect of $b(k)$ is to reduce the residue of the pole of G from 1 to $1/(1+ib)$ and to decrease the damping from $\nu = a$ to $\nu = a/(1+b^2)$. The correlation function is reduced by the same amount:

$$\hat{U}(k, 0) = \int \frac{d\omega}{2\pi} U(k, \omega) = \frac{F_0(k) + \Sigma_2(k)}{2\nu(k)(1+b^2(k))}.$$

We wish to check the size of b numerically. We take $\hat{U}(k, 0)$ to be given by GM76, a recently revised version of the Garrett-Munk spectrum (see McComas & Bretherton

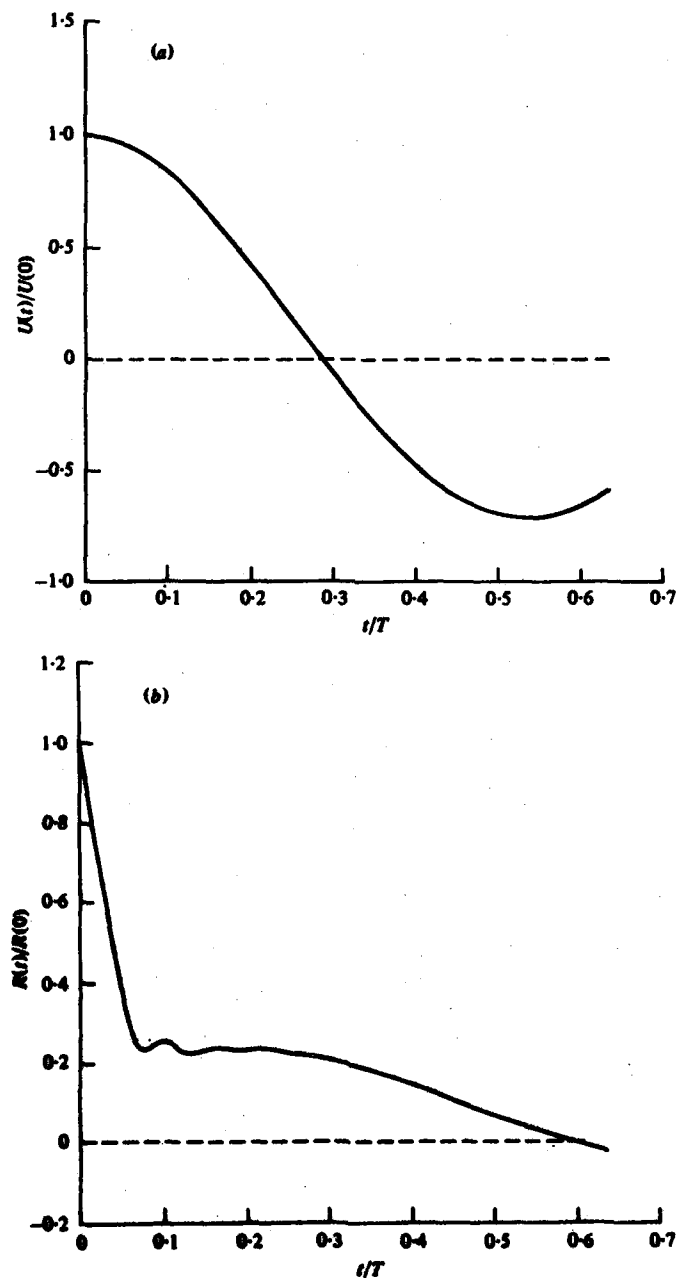


FIGURE 10. (a) $U(t)/U(0)$ vs. t/T for $\lambda_V = 300$ m and $\Omega/f = 4.5$ and (b) $R(t)/R(0)$ vs. t/T for $\lambda_V = 300$ m and $\Omega/f = 4.5$.

1977). The coupling coefficients used in the calculation are derived in the same manner as those given by McComas.

We have computed values of $a(k)$ and $b(k)$ by assuming that both were initially zero (i.e. just $\Sigma_1 \rightarrow 0$ as before), and then iterating (3.7) once. Values of $a(k)/\Omega(k)$ and $b^2(k)$ are shown in figures 8 and 9 respectively. We first note the previously mentioned fact that the values of $a(k)$ for much of the spectrum violate

$$a(k)/(\Omega(k)/2\pi) \leq 1.$$

The dashed line corresponds to $a(k)/(\Omega(k)/2\pi) = 1$. The RIA assumes that $\Omega > a$; that condition is clearly violated here. We further note that an examination of figure 9 reveals values of $b^2(k)$ that are extremely large for much of the spectrum. This clearly violates the assumption that Σ_1 is constant in ω . We conclude that any reliable self-consistent calculation must include information on the non-resonant structure of Σ_1 . It should be noted, however, that the simple linear form of Σ_1 , chosen here to test the constant $-\Sigma_1$ assumption, is wholly inappropriate for a serious calculation.

Finally, we can also determine the validity of the assumption that Σ_2 is constant in ω . When we actually compute $\Sigma_2(k, \omega)$ we find that Σ_2 actually does possess an appreciable structure with a finite width in ω . However it is not sufficient merely to calculate Σ_2 for various values of ω , since frequency filter cut-off may allow only values of Σ_2 from a narrow band around resonance to enter the calculation. In that case Σ_2 may be considered *effectively* constant. One way to circumvent this is to compare the functions

$$U(t) = \int d\omega e^{i\omega t} \frac{\Sigma_2(\omega)}{\nu^2 + (\omega - \Omega)^2}$$

and

$$R(t) = \int d\omega e^{i\omega t} \Sigma_2(\omega).$$

We note that Σ_2 always appears in the equations through $U(\omega)$, which contains information on the width of the frequency filter as well as on the width of Σ_2 . On the other hand $R(t)$ contains information on the width of Σ_2 only. We envision two possible extreme cases. In one case ν is extremely small, so that the frequency filter becomes a resonance delta function and $U(t)$ assumes the form earlier obtained by assuming $\Sigma_2 = \text{constant}$. On the other extreme, ν might be so large that the width of the frequency filter would be much larger than the width of Σ_2 . In this case we would expect the shapes of $U(t)$ and $R(t)$ to be identical. The calculations that we have performed indicate that both extreme situations do indeed obtain in different regions of the spectrum. Figure 10 shows the first case. Here Σ_2 is constant and $U(t) \propto e^{i\Omega t} e^{-\nu|t|}$. If Σ_2 were actually constant, however, $R(t)$ should be a delta function in t . In actuality we see that $R(t)$ does have some width, indicating that Σ_2 is not constant but that the variation is unimportant in the calculation. On the other hand figure 11 shows the case where, appropriately scaled, $U(t)$ and $R(t)$ are identical, indicating large values of $\nu(k)$. In this region, then, not only is the RIA very bad, but also the assumption that $\Sigma_2 = \text{constant}$ must be dropped if the frequency filter correction is to be used.

We have also calculated $\Sigma_1(\omega)$ and from it the response function to verify directly that the frequency dependence of Σ_1 is important for some modes.

In general, then, we conclude that any self-consistent calculation of Σ_1 and Σ_2 in the region where Oibers and McComas found inconsistencies in their calculations must necessarily include (1) a frequency filter function to relax the RIA, and (2) non-resonant

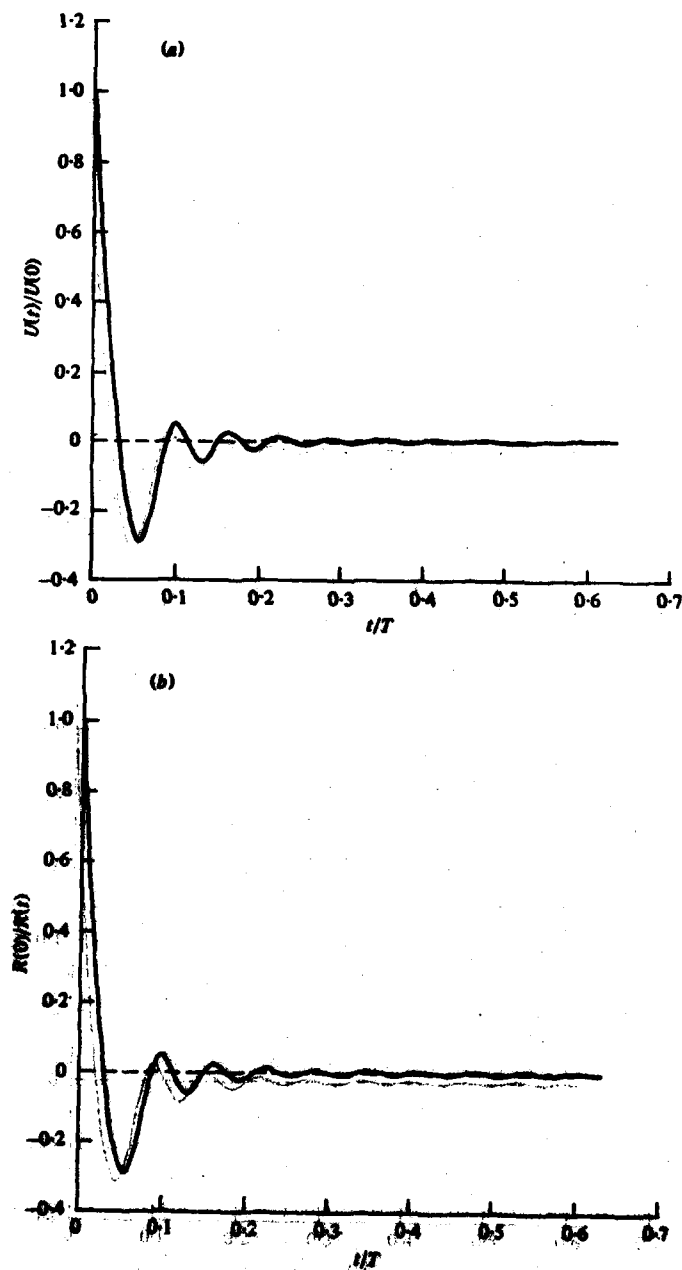


FIGURE 11. (a) $U(1)/U(0)$ vs. $1/T$ for $\lambda r = 1.5$ m and $Q/U = 4.5$ and (b) $R(1)/R(0)$ vs. $1/T$ for $\lambda r = 1.5$ m and $Q/U = 4.5$.

information on Σ_1 and Σ_2 . Unfortunately this greatly increases the difficulty of the numerical solutions of the problem, since it requires an extra integration over ω , including a principal part integral that has not been present in earlier calculations. In addition, the equations must be solved self-consistently, implying an iteration process on an already time-consuming calculation. We are currently attempting a simplified version of the self-consistent calculation.

6. WKB approximation

In the introduction we alluded to other approximations to the path integral. One of the failings of weak interaction theory can be traced to the interaction of two small-scale high-frequency waves with a large-scale inertial frequency wave. This interaction has been called induced diffusion by McComas & Bretherton (1977). It is appropriate to think of it as the effect of advection of the small-scale wave by the inertial wave. This picture suggests studying the propagation of a small-scale wave packet in the geometric optics approximation including the effects of a time-varying background. The aim would be to calculate a coherent decay time. The time scales of the response functions in the induced diffusion region are dominated by processes that leave the wave packet coherent, although somewhat altered.

7. Summary

We have written this paper with a twofold purpose. In the first part of the paper we have introduced a formalism for solving nonlinear equations of the type encountered with internal waves. From experience with self-consistent field treatments in other areas of physics we very well might expect such methods to be helpful in extending calculations beyond some of their inconsistencies. To this end we have taken a simplified steady-state model of the oceanic internal wave field and obtained via path integral methods the corresponding diagrammatic perturbation theory from which Dyson's equations can be obtained. In the process we have suggested as interpretation of the ocean as a collection of uncoupled, damped oscillators driven by random forces and each having a memory. In this case we used the direct interaction approximation of Kraichnan, which is easily implemented in the diagrammatic language. We should note that one problem with this approach is that it is not currently known what kind of approximations guarantee the positivity of quantities like the effective viscosity and also guarantee consistency in the Green's function sum rule. In turbulence problems it has been possible to construct models of fictitious systems that satisfy the DIA exactly, thus guaranteeing the appropriate positivity. However this approach is of very limited usefulness and it is not known how to extend it to other approximations.

In the second part of the paper we have obtained the previous results of Hasselmann via a prescribed limiting process. The particular assumptions inherent in the process were checked and found to break down in the region where Olbers and McComas found inconsistent results in their calculations. We suggested that the calculations might be improved by solving equations for Σ_1 and Σ_2 self-consistently, but that the implementation of the procedure might be quite difficult.

In conclusion, we feel that the formalism of the type discussed here should be

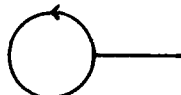


FIGURE 12. Tadpole graph with response function integrated.

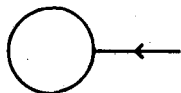


FIGURE 13. Tadpole graph with correlation function integrated.

studied further for its possible application to the internal wave problem. The preliminary calculations indicate ways to correct for deficiencies in earlier calculations, but further work is needed to formulate positively constraints and related constraints.

This work was supported in part by the Navy SBC, Johns Hopkins University, APL 601103, and by the Office of Naval Research, N00014-80-C-0840.

Appendix

We discuss briefly the role of the Jacobian (equation (2.14)) in cancelling some tadpole graphs. We will illustrate the cancellation by considering the expectation value of $A(\mathbf{k}, \omega)$, which should be zero. There are two graphs contributing to $\langle A(\mathbf{k}, \omega) \rangle$ and they are shown in figures 12 and 13. If we work to lowest order in B , the Jacobian gives the contribution

$$\langle A(\mathbf{k}, \omega) \rangle_{\text{Jacobian}} = 2 \int d^3\mathbf{k}_1 d\omega_1 \sum_{s_1 s_2} \frac{\langle A^s(\mathbf{k}_1 = 0, \omega_1 = 0) A^s(\mathbf{k}, \omega) \rangle_0}{D^s(\mathbf{k}_1, \omega_1)} B_{\mathbf{k}_1 - \mathbf{k}, 0}^{s_1 - s_2}, \quad (\text{A } 1)$$

where $\langle AA \rangle_0$ is given by (3.1). Using the rules given in the text it is easy to calculate the contribution of figure 12 and verify that it cancels (A 1). The graph shown in figure 13 vanishes for another reason. It is easy to see that it is proportional to

$$\frac{1}{D^s(\mathbf{k} = 0, \omega = 0)} \sum_{s_1} \int d^3\mathbf{k}_1 d\omega_1 B_{\mathbf{k}_1 - \mathbf{k}, 0}^{s_1 - s} U_0^s(\mathbf{k}_1, \omega_1)$$

and

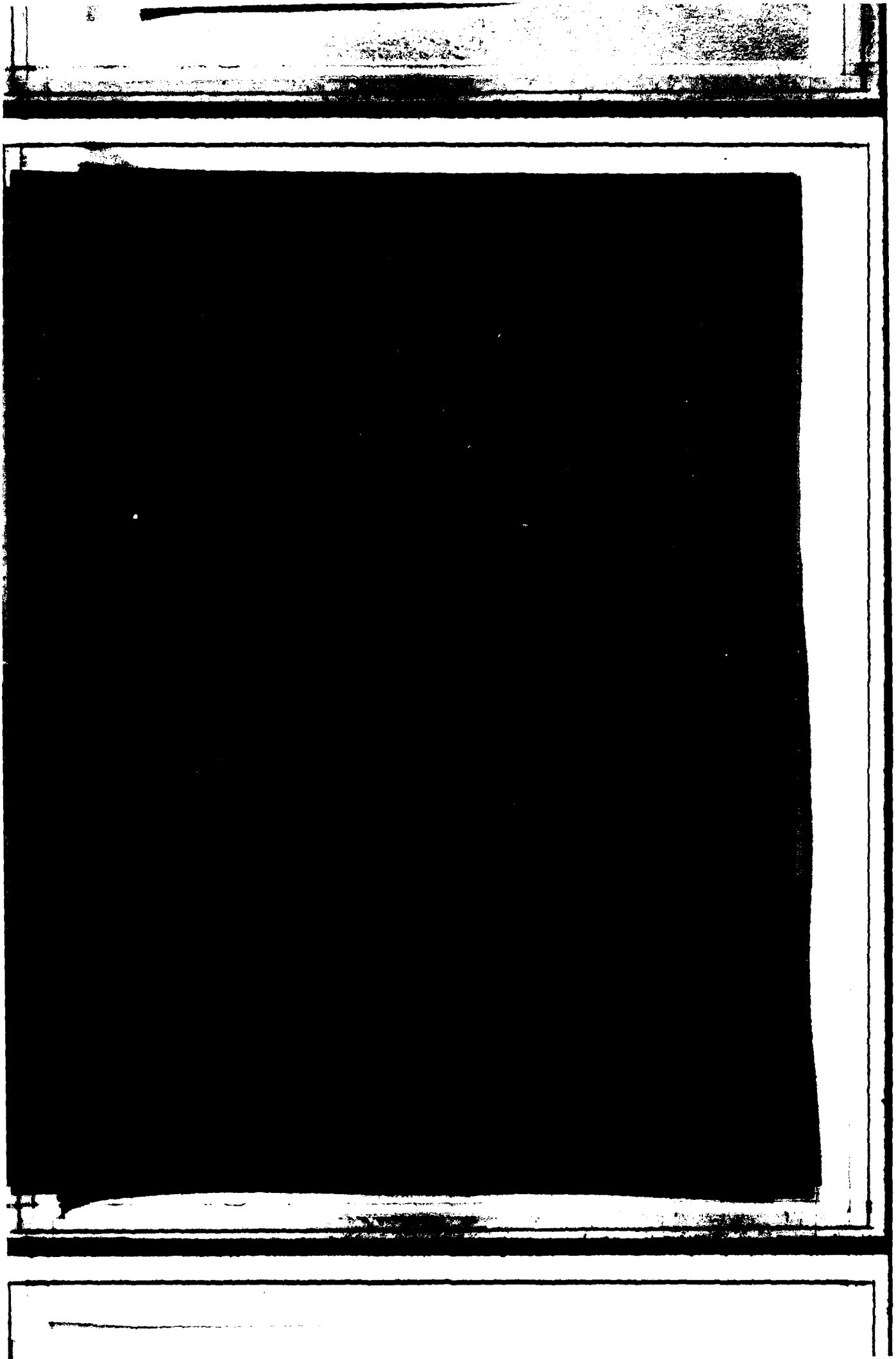
$$B_{000}^{s-s} = 0.$$

When these graphs are inserted into more complicated graphs they will continue to give no contribution. Similarly more complicated tadpoles (vertex corrections of figures 12 and 13) will continue to vanish.

REFERENCES

- EDWARDS, S. F. 1964 The statistical dynamics of homogeneous turbulence. *J. Fluid Mech.* **18**, 239-273.
 GARRETT, C. J. R. & MUNK, W. H. 1972 Space-time scales of internal waves. *Geophys. Fluid Dyn.* **2**, 235-264.
 GARRETT, C. J. R. & MUNK, W. H. 1975 Space-time scales of internal waves: A progress report. *J. Geophys. Res.* **80**, 291-297.

- HASSELMAIR, K. 1966 Feynman diagrams and interaction rules of wave-wave scattering processes. *Rev. Geophys. Space Phys.* **4**, 1-32.
- HASSELMAIR, K. 1967 Nonlinear interactions treated by the methods of theoretical physics. *Proc. R. Soc. Lond. A* **292**, 71-100.
- HASSELMAIR, K. 1968 Weak interaction theory at ocean waves. In *Basic Developments in Fluid Dynamics*, vol. 2, pp. 117-132.
- KHARIN, J. R. 1965 Self-consistent-field approach to turbulence theory. *Phys. Fluids* **8**, 2319-2323.
- HOLLOWAY, G. 1970 Order and disorder in nonlinear fluid motions. *Topics in Nonlinear Dynamics - A Tribute to Sir Edward Bullard*. American Institute of Physics.
- HOLLOWAY, G. 1970 On the spectral evolution of strongly interacting waves. *Geophys. Astrophys. Fluid Dyn.* **11**, 271-287.
- HOLLOWAY, G. & HANDELMAN, M. C. 1977 Stochastic closure for nonlinear Rossby waves. *J. Fluid Mech.* **82**, 747-765.
- KRAICHNAN, R. H. 1969 The structure of isotropic turbulence at very high Reynolds numbers. *J. Fluid Mech.* **5**, 599-615.
- LANGMUIR, I., NEUFELDER, D. & TATARU, E. 1970 Functional integral methods for stochastic fields. *Physica A* **55**, 232.
- LEWIS, D. G. 1972 *Developments in the Theory of Turbulence*. Oxford University Press.
- McCORMACK, C. H. 1977 Equilibrium mechanisms within the oceanic internal wave field. *J. Phys. Geophys.* **7**, 555-565.
- McCORMACK, C. H. & BARNETT, P. P. 1977 Resonant interactions of oceanic internal waves. *J. Geophys. Res.* **82**, 1401-1412.
- MANN, F. C., SUGA, E. D. & ROSS, R. A. 1975 Statistical dynamics of classical systems. *Phys. Rev. A* **12**, 123-137.
- OSWALT, R. 1970 Stochastic energy transfer and the energy balance of the internal wave field in the deep ocean. *J. Fluid Mech.* **44**, 505-520.
- OSWALT, R. A. 1977 Lecture on the statistical theory of turbulence. In *Fluid Dynamics* (ed. R. Bullard & J. L. Parker), p. 235. Gordon and Breach.
- PELLING, C. H. 1977 *The Dynamics of the Upper Ocean*. Cambridge University Press.
- PRESTER, R. 1968 Statistical mechanical models for stationary homogeneous turbulence. *J. Phys. A* **1**, 121-132.
- PRESTER, R. 1977 The functional formalism of classical statistical dynamics. *J. Phys. A* **10**, 771-785.
- FOURMANT, N., MONT, J. D. & WATSON, K. M. 1970 Spectral analysis of nonlinear internal wave interactions using Langmuir methods. *J. Geophys. Res.* **75**, 4935-4945.
- WYLLIE, H. W. 1961 Formulation of the theory of turbulence in an incompressible fluid. *Ann. Phys.* **14**, 143-165.



ABSTRACT

In this paper we apply a formalism introduced in a previous paper to write down a self-consistent set of equations for the functions that describe the near-equilibrium time behavior of random oceanic internal waves. These equations are based on the Direct Interaction Approximation. The self-consistent equations are solved numerically (using the Garrett-Munk spectrum as input) and the results are compared to parameters obtained in the weak interaction approximation (WIA). The formalism points out that an extra parameter that is implicitly vanishingly small in the WIA has a significant effect on decay rates when computed self-consistently. We end by mentioning possible future self-consistent calculations that would improve upon our own.

1. INTRODUCTION

In the past few years the study of transfer of energy in the internal wave field of the ocean has become an active area of research. Olbers (1976), McComas (1977), Pomphrey, Meiss and Watson (1980), and DeWitt and Wright (1982), have done calculations in the resonant interaction approximation (RIA) of lifetimes and action transfer rates.

These calculations assumed that the interactions were weak and the lifetimes of particular modes were long compared to the intrinsic frequencies. Because of the diversity of scales in the ocean, this approximation is invalid. The large scale, near inertial frequency waves interact strongly with the smaller scale waves. Attempts at solving this part of the problem have been made by Meiss and Watson (1982) and Hanyey and Pomphrey (1982). In this paper we present a calculation using the formalism presented in DeWitt and Wright (1982) and DeWitt (1982). The idea is to use the Direct Interaction Approximation (DIA). In fact we will use additional approximations to the DIA. The details and further references can be found in DeWitt and Wright (1982) and DeWitt (1982).

The DIA equations give an effective linear theory with memory terms and forcing terms. The memory function and forcing function are determined by a self consistent calculation that will be described in Section II. In this

section we write down the effective linear theory as that will enable the reader to better understand what is being calculated.

Our formulation of the nonlinear problem is the same as that of McComas (1977) or Meiss, Pomphrey and Watson (1979). We very briefly review the equations. A Lagrangian is introduced with the variable being the displacement field $\xi_j(r, t)$

$$\begin{aligned}
 L(\xi) = & \frac{1}{2} \rho (\dot{\xi}_j \dot{\xi}_j + \epsilon_{jkl} f_j \dot{\xi}_k \xi_l) \\
 & - \rho g \xi_j \delta_{j3} - \left[\xi_j \frac{\partial p_e}{\partial r_j} + \frac{1}{2} \xi_j \xi_k \frac{\partial^2 p_e}{\partial r_j \partial r_k} + \dots \right] \\
 & - \left(\xi_j \frac{\partial \pi}{\partial r_j} + \frac{1}{2} \xi_j \xi_k \frac{\partial^2 \pi}{\partial r_j \partial r_k} + \dots \right)
 \end{aligned} \tag{1.1}$$

ξ is then expanded in terms of the normal modes

$$\xi_j(\vec{r}, t) = \sum_{s=\pm} \int d^3k A_k^s(t) Z_{kj}^s e^{i\phi_k^s} \tag{1.2}$$

$A(t)$ is the mode amplitude and $\phi_k^\pm = \vec{k} \cdot \vec{r} \mp \Omega_k t$

Ω_k is the dispersion relation,

$$\Omega_k = \sqrt{\frac{N^2 K_H^2 + f^2 K_V^2}{K_H^2 + K_V^2}}$$

The equations of motion for the amplitudes A are given in the next section. For details on the coefficients Z and the couplings B of the next section see DeWitt (1982) or Olbers (1976). In the DIA approximation all of the Greens functions and correlation functions are identical to those of the following linear system of equations:

$$\begin{aligned} \frac{dA^-}{dt}(k,t) + \int_{-\infty}^{\infty} \left(\Gamma_{+-}(k,t,t') A^-(k,t') \right. \\ \left. + \Gamma_{++}(k,t,t') A^+(k,t') \right) dt' = f^-(k,t) \\ \frac{dA^+}{dt}(k,t) + \int_{-\infty}^{\infty} \left(\Gamma_{--}(k,t,t') A^-(k,t') + \Gamma_{-+}(k,t,t') A^+(k,t') \right) dt' \\ = f^+(k,t) \end{aligned} \quad (1.3)$$

$\Gamma(t,t')$ is a function determined by the DIA equations. Casuality demands $\Gamma(t,t') = 0$ for $t' > t$. We are treating a stationary problem so Γ depends only on $t-t'$. $f(k,t)$ is a random forcing function on the k^{th} mode due to all of the other modes and any external random forces. In the DIA approximation the f 's are Gaussian random variables with zero mean (provided that the external forces are also Gaussian with zero mean). Their correlation function is given by

$$F_0^s(k,t-t') \delta_{s,-s'} + \Sigma_2^s s' (k,t-t') = \langle f^s(k,t) f^{s'}(k',t') \rangle \quad (1.4)$$

F_0 is the correlation function for the external part of the forces (1.4) and Σ_2 is calculated from the DIA equations. It is convenient to work with the

Fourier transform in the time domain, then Γ has the following representation,

$$\Gamma^{ss'}(k, \omega) = \delta_{s, -s'} \left[v_0(k) + i \Omega^{-s}(k) \right] - \Sigma_1^{ss'}(k, \omega) \quad (1.5)$$

Here v_0 is the damping due to interaction with the external environment and $\Omega^\pm = \pm \Omega_k$. If Σ_1 were independent of ω it would represent three effects. The obvious ones are a finite lifetime and frequency shift. In addition however, there is a coupling between the \pm modes. Previous calculations have used weak interaction theory which is the limit of the DIA equation in which Σ_1 and Σ_2 are independent of ω and infinitesimal.

In the following section we introduce the correlation function for the A's

$$\frac{1}{2\pi} U^{ss'}(k, \omega) \delta^3(\vec{k} + \vec{k}') \delta(\omega + \omega') = \langle A_k^s(\omega) A_{k'}^{s'}(\omega') \rangle \quad (1.6)$$

and the Green's function for equations (1.3) ,

$$G^{ss'}(k, \omega) = \left[v_0 + i \Omega_k^s - i \omega - i \Sigma_1^{ss'}(k, \omega) \right]^{-1} \quad (1.7)$$

The inverse is a matrix inverse of a 2×2 matrix.

The only additional information that one gets from the DIA equations other than the correlation and Green's functions comes from an inspection of the calculation of Σ_1 and Σ_2 where it is possible to deduce where the action in a mode k is flowing and where it came from. We do not present any of those results in this paper. From the fact the DIA equations give correlation and Green's functions that are equivalent to a linear theory in which different

modes are uncoupled, it is clear that they can't give a completely accurate picture of the internal wavefield. The effect of these correlations is lumped into finite lifetime and frequency shifts. Thus we can expect some improvement on weak interaction theory, but the interaction between diverse scales may still not be adequately treated.

We find substantial differences from weak interaction rates for many of the internal wave modes. That is the functions $\Sigma_1(k)$ and $\Sigma_2(k)$ are quite different from those of weak interactions. One problem that immediately arises in this case is the validity of truncating the Lagrangian in Eq. 1.1 at third order. An Eulerian formulation would remove this problem but introduce others. In this paper we use Eq. (1.1) and ignore higher order corrections without attempting to justify that approximation.

The remainder of the paper is organized into four sections. In Section II we write down the DIA equations and our approximation to them. In Section III we discuss briefly the meaning of the parameters and in Section IV we give the results of our calculations. Section V gives our conclusions.

2. THE SELF-CONSISTENT EQUATIONS

In this section we write down the specific equations that will be used in the numerical solution of the problem. The derivation of the more general Direct Interaction Approximation (DIA) equations is given in an earlier paper. (Dewitt and Wright, 1982). The exact DIA equations are

$$\Sigma_1^{ss'}(k, \omega) = \frac{2}{\pi} \sum_{\substack{s_1, s_1' \\ s_2, s_2'}} \int d^3k_1 d^3k_2 d\omega_1 d\omega_2 \times$$

$$\delta(k+k_2-k_1) \delta(\omega+\omega_2-\omega_1) B_{k_1 k_2-k_1}^{-s_1 s_2 -s_1} \times \quad (2.1)$$

$$B_{k_1-k_2-k}^{-s_1' s_2' -s_1'} [iG^{s_1 s_1'}(k_1, \omega_1)] [-U^{s_2 s_2'}(k_2, \omega_2)]$$

and

$$\Sigma_2^{ss'}(k, \omega) = \frac{1}{\pi} \sum_{\substack{s_1, s_1' \\ s_2, s_2'}} \int d^3k_1 d^3k_2 d\omega_1 d\omega_2 \times$$

$$\delta(k+k_2+k_1) \delta(\omega+\omega_2+\omega_1) B_{k_1 k_1 k_2}^{-s_1 s_1 s_2} \times \quad (2.2)$$

$$B_{-k k_1 -k_2}^{-s_1' s_1' s_2'} U^{s_1 s_1'}(k_1, \omega_1) U^{s_2 s_2'}(k_2, \omega_2)$$

where G is given by Eq. 1.7 and U by

$$U^{ss'}(k, \omega) = G^{ss_1}(k, \omega) [\delta_{s, -s'} F_0(k, \omega) + \Sigma_2^{s_1 s_2}(k, \omega)] G^{s' s_2}(-k, -\omega]$$

All of the theoretical information about the interaction of the internal waves is contained in the coupling coefficients $B_{k_1 k_2 k_3}^{s_1 s_2 s_3}$ which in turn are derived from first principles and include some approximations. They are dynamically defined by the equations of motion for the internal wave amplitudes A :

$$\dot{A}_k^s(t) + i\Omega_k^s(t) + v_0(k)A_k^s(t) + \sum_{1,2} \int d^3k_1 d^3k_2 \varepsilon(k-k_1-k_2) \times$$

$$B_{k-k_1-k_2}^{s-s_1-s_2} A_{k_1}^{s_1} A_{k_2}^{s_2} = f_0^s(k, t) \quad (2.4)$$

The goal of the calculation is to compute the matrices Σ_1 and Σ_2 in a self-consistent manner. All of the information about the decay and correlation of the waves is contained in these two quantities.

We hasten to remind the reader that even a complete solution of the DIA equations does not constitute an exact solution of the problem, since these equations were obtained by completely ignoring all vertex corrections in the

relevant diagrams. It is obviously not clear a priori that such an approach is reasonable; one must eventually compute at least the leading terms in the vertex expansions to vindicate the procedure.

As was pointed out in the previous paper, much has been said about possible ways to solve systems of equations that are related to the full DIA set of equations. However the only numerical attempts at solution in this framework have used the Weak Interaction Approximation (WIA), which computes Σ_1^{+} and Σ_2^{+} by letting

$$\Sigma_1, \Sigma_2 \rightarrow 0 \quad (2.5)$$

in the expressions on the right hand side of equations 2.1 and 2.2. The only outcome that would justify this approach would be one in which the computed values of Σ_1 were small compared with typical oscillation frequencies. As has been pointed out many times before, this is not the case. This makes the WIA a non-self-consistent approach.

Our parameterization of the DIA equations assumes that $G(\vec{k}, \omega)$ can be represented by two poles, one at $\omega = \Omega_r - i\alpha_I$ and a second at $\omega = -\Omega_r - i\alpha_I$. Symmetry between $\pm \vec{k}$ is assumed as well as between $\pm \Omega$. The equations are then parameterized by the pole positions. Also we require $G^{ss'}(\vec{k}, t = 0^+) = \delta_{s, -s'}$. Finally, there is the question of where to evaluate $\Sigma_1(\omega)$. (Carnevale and Fredericksen, 1982). That is do we evaluate it at the pole for complex ω or somewhere on the real axis. We chose to evaluate $\Sigma_1(\omega)$ at $\omega = \Omega(k)$ with Ω given by the linear dispersion relation. See Section 3 for further comments. In particular we choose

$$\Sigma_1(k, \omega) = \Sigma_1(k, \Omega)$$

$$\Sigma_2(k, \omega) = \Sigma_2(k, \Omega) \quad (2.6)$$

The quantities Σ_1 and Σ_2 are in general 2X2 matrices with complex components and so superficially appear to represent 16 real parameters. However using the constant- ω assumption and the general symmetry conditions

$$\begin{aligned} U^{ss'}(k, \omega) &= [U^{-s, -s'}(-k, -\omega)]^* \\ U^{ss'}(k, \omega) &= U^{s's}(-k, -\omega) \\ G^{ss'}(k, \omega) &= [G^{-s, -s'}(-k, -\omega)]^* \end{aligned} \quad (2.7)$$

enables us to reduce these to 7 independent real quantities. We denote

$$\begin{aligned} a(k) &\equiv \Sigma_1^{-+}(k, \Omega) \\ c(k) &\equiv \Sigma_1^{++}(k, \Omega) \\ d(k) &\equiv \Sigma_2^{++}(k, \Omega) \\ e(k) &\equiv \Sigma_2^{-+}(k, \Omega) \end{aligned} \quad (2.8)$$

The first symmetry condition implies that $e(k)$ is real; the other three quantities are in general complex.

By invoking a further assumption we can reduce this set of parameters even further. We proceed as follows: with the above assumption we can write

the Green's function for this problem as

$$G(k, \omega) = \frac{1}{(\omega - \eta)(\omega - \sigma)} \begin{pmatrix} c^* & \omega + \Omega - a^* \\ \omega - \Omega + a & -c \end{pmatrix} \quad (2.9)$$

In writing matrices we will use the convention that $s = +$ is represented by the first row (or column). In the expression for G

$$\begin{aligned} \eta &\equiv \Omega_r - ia_I \\ \sigma &\equiv -\Omega_r - ia_I \\ \Omega_r^2 &\equiv (\Omega - a_R)^2 - |c|^2 \end{aligned} \quad (2.10)$$

Here Ω_r is the renormalized "frequency"; a_R and a_I represent the real and imaginary parts of a respectively. We note that in the weak interaction limit $\Omega_r \rightarrow \Omega$, but that in this self-consistent calculation Ω_r may be either purely real or purely imaginary depending on the relative values of $(\Omega - a_R)$ and $|c|$. This will be explored more fully later.

Notice that in this particular case the two poles of the Green's function satisfy

$$\begin{aligned} (\eta + \sigma)^* &= -(\eta + \sigma) \\ (\eta \sigma)^* &= \eta \sigma \end{aligned} \quad (2.11)$$

This result can be seen directly by comparing with 2.10, but more importantly it can be obtained by applying the symmetry property of G in 2.7 without knowing the analytical expressions for η and σ . This is useful because the same structure appears when Green's functions with more poles are considered.

In such cases it is not generally possible to obtain closed form expressions for the poles.

From the equation for G we can compute the correlation function

$$U(\omega) = G(\omega) \Sigma_2 G^T(-\omega) \quad (2.12)$$

where the superscript T denotes transpose of the matrix. If we denote

$$U(\omega) = [(\omega-\eta) (\omega+\eta) (\omega-\sigma) (\omega+\sigma)]^{-1} \tilde{U}(\omega) \quad (2.13)$$

then $\tilde{U}(\omega)$ has no poles.

Now we are ready to impose "experimental" constraints on our expressions.

The function

$$U(k,t) = \int d\omega e^{i\omega t} U(k,\omega) \quad (2.14)$$

Gives the two-time correlation of the internal wave amplitudes and so the quantity

$$U^{ss'}(t=0) = \mathcal{W}(\vec{k}_H, k_V) \delta_{s,-s'} \quad (2.15)$$

is given experimentally by the Garrett-Munk spectrum (Garrett and Munk 1972, 1975). We take this as an additional constraint on the equations. If the expressions for U are made to satisfy this constraint it is found that

$$\begin{aligned} d &= -ic (e/a_I) \\ \mathcal{W} &= e/2a_I \end{aligned} \quad (2.16)$$

This last expression is the same one obtained in the WIA. We see that Σ_2 is completely determined by Σ_1 and the Garrett-Munk spectrum. This leaves us

with only four independent real quantities to compute. The expressions for \tilde{U} under this approach are also greatly simplified:

$$\tilde{U} = e \begin{pmatrix} \frac{ic^*}{a_I} (\omega^2 + \eta\sigma) & (\omega + \Omega - a_R)^2 + a_I^2 - |c|^2 \\ (\omega - \Omega + a_R)^2 + a_I^2 - |c|^2 & \frac{-ic}{a_I} (\omega^2 + \eta\sigma) \end{pmatrix} \quad (2.17)$$

The great benefit of this approach is that we now need only iterate one of the two DIA equations (the equation giving Σ_1). As a matter of fact if one reviews the steps taken to derive the final expression for G and U it is not difficult to see that this is generally true whenever Σ_2 is chosen to be constant, regardless of the form of Σ_1 . After the Σ_1 equation has been solved self-consistently the value of e may be computed from the Σ_2 equation to determine how good the ratio $e/2a_I$ actually gives back the Garret-Munk spectrum. If we use the equations in Section 1, it is easy to derive the following balance equation:

$$\frac{F_0(k)}{\mathcal{U}(k)} - 2\nu_0(k) = 2\text{Im } \Sigma_1(k) - \frac{\Sigma_2(k)}{\mathcal{U}(k)} \quad (2.18)$$

If there is no forcing or dissipation on mode k due to noninternal wave sources, the left hand side is zero. In this case if the right hand side is nonzero one would infer some inconsistency. One possibility is that the Garrett-Munk spectrum $\mathcal{U}(k)$ is not correct. A second is of course that the theoretical calculation of Σ_1 and Σ_2 is invalid. The third possibility is that there is either an effective F_0 present or an effective ν_0 , i.e. mode k might be driven or damped by direct coupling to the external environment.

3. PHYSICAL SIGNIFICANCE OF THE PARAMETERS

Before we display our numerical results for the parameters Σ_1 and Σ_2 we would first like to discuss the meaning of the parameters introduced in the last section. To do this we recall that the nonlinear evolution Equation 2.4 may be equivalently written as a set of effective linear evolution equations, which for the constant Σ case considered in this paper are

$$\begin{aligned} \dot{A}^+(k,t) + (\nu_0 + i\Omega - ia) A^+(k,t) + iC^* A^-(k,t) \\ = f^+(k,t) \\ \dot{A}^-(k,t) + (\nu_0 - i\Omega + ia^*) A^-(k,t) - iCA^+(k,t) \\ = f^-(k,t) \end{aligned} \tag{3.1}$$

The functions f^+ and f^- give the effective driving of the waves due to all other modes. The parameter ν_0 is a free parameter that may be used to inject additional information about the decay of internal waves due to interaction with external systems. In all numerical work we always set $\nu_0 = 0$.

The parameter a represents two physical quantities. The imaginary part of a represents an additional effective damping of the internal wave mode due to losses to other internal waves modes. The real part of a gives the frequency shift of the wave. The parameter a is not new to this type of

analysis. The parameter c , however, has not to our knowledge appeared in this context before. We see that the formalism compels us to include a parameter that represents a coupling between $+$ and $-$ traveling waves.

The pair of first order equations for the amplitudes may be written as a pair of second order equations involving only one type of amplitude each. If we ignore the f , then for either A^+ or A^- we have

$$\ddot{A}(k,t) + 2(\nu_0 + a_I) \dot{A}(k,t) + [(\Omega - a_R)^2 + (\nu_0 + a_I)^2 - |c|^2] A(k,t) = 0 \quad (3.2)$$

The solution has a time dependence

$$A^\pm \sim \exp \left\{ -(\nu_0 + a_I) t \mp \sqrt{|c|^2 - (\Omega - a_R)^2} t \right\} \quad (3.3)$$

If $|c|^2 > (\Omega - a_R)^2$, then we refer to the mode as overdamped. In the underdamped case the damping is given by

$$\nu = \nu_0 + a_I \quad (3.4)$$

In the overdamped case the damping is

$$\nu = \nu_0 + a_I \pm |\Omega_R| \quad (3.5)$$

We will define

$$\begin{aligned}\alpha_s &= a_I + |\Omega_r| \\ \alpha_L &= a_I - |\Omega_r|\end{aligned}\tag{3.9}$$

In the overdamped region

$$\begin{aligned}A^+ &\sim e^{-\alpha_s t} \\ \text{and} \\ A^- &\sim e^{-\alpha_L t}\end{aligned}\tag{3.7}$$

A spike disturbance in the equilibrium spectrum will contain both a short lived and a long lived component. We will see from the results of the numerical calculations that it is α_L that measures the lifetime of a disturbance for a large part of the internal wave spectrum.

4. NUMERICAL RESULTS

In Figure (1) we show a contour plot of $\log |c/(\Omega - a_R)|$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis. Here m refers to the vertical wavenumber. The line of critical damping has the value zero; everything to the left of this line is in the underdamped regime, while everything to the right is in the overdamped regime. It is instructive to compare this plot with Figure (2), a plot of $\log(a_I(WI)/\Omega)$ for the weak interaction (WI) case using the same wavevector cutoffs for both. Notice that the SC overdamped region covers the entire $a_I/\Omega > 1$ region and overlaps the $a_I/\Omega < 1$ region, so that the overdamping takes over just at the place where the old weak interaction calculation begins to break down. What is interesting is the extent to which the self-consistent (SC) calculation reduces the rates in the large- m , large- Ω region. Of course we do expect the SC calculation to reduce the damping rates in this region, for reasons described below, but the way that this reduction comes about is rather unexpected. Recall that the rate of decay of a spike disturbance introduced into the steady state ocean is given by α_L and α_S in the SC calculation. (See Eq.(36)) If we plot $\log|\alpha_L/\alpha_S|$ in Figure (3) we see that in essentially the entire overdamped region $\alpha_L \ll \alpha_S$, which implies that $a_I(SC) \approx |\Omega_R|$. This means that the decay rates are much smaller than one would expect by examining the

scale of a_I . Unfortunately the same device that causes the damping rates to be so greatly diminished also makes it more difficult to obtain accurate numbers in the large- m region. A small change in the parameters (a,c) in this region can cause a disproportionately large change in α_L (though not in α_S). Since it is (a,c) that are computed in each iteration, factors such as the integration accuracy make it difficult to control such changes after a certain limit is reached. We have solved the equations by an iterative scheme. After many iterations the results do not change much, but the cancellations and inaccuracies in the integration scheme lead to some changes from iteration to iteration. In the next two figures we compare the a_I from two successive iterations. Figure (4) is a plot of the percent deviation of the input values of a_I compared to the output values of a_I .

$$\Delta(a_I) \equiv 100 \left| \frac{a_I(\text{in}) - a_I(\text{out})}{a_I(\text{in}) + a_I(\text{out})} \right| \quad (4.1)$$

In a perfect calculation, of course, these numbers would be identical. We see in fact that in most of the region of interest we have converged to within a few percent, and it is only for very large values of m that we reach values of 8% to 10%. However, if we compare this with Figure (5), which is a plot of

$$\log \Delta(\alpha_L) \equiv \log 100 \left| \frac{\alpha_L(\text{in}) - \alpha_L(\text{out})}{\alpha_L(\text{in}) + \alpha_L(\text{out})} \right| \quad (4.2)$$

we see that the error increases in α_L more quickly than it does in a_I . In the large- m region this means that we only know α_L within a factor of 2 or so. We hasten to point out that this does not alter our conclusions about the dramatic effects of the SC calculation. First of all the fact that the values of (a,c) for a given mode do not depend on the behaviour of modes with even

moderately greater values of (m, Ω) means that the computed values in most of the space are not affected by the uncertainty in the large- m extreme. Second, we will see shortly that compared to the WIA the rates for the SC case in the large- m region are reduced by orders of magnitude, and so even a factor of 2 uncertainty does not eliminate the qualitatively new behavior.

Referring back to (3.7) we see that $\alpha_L \ll \alpha_S$ implies that, for example,

$$A^-(k, t) \sim e^{-\alpha_L t} \quad (4.3)$$

so that we may ignore α_S and consider only α_L in determining the long time behavior of G . For this reason we plot $\log(\alpha_L/\Omega)$ in Figure (6) with the understanding that $\alpha_L = a_L$ in the underdamped regime. It is clear that these rates are very much different from those calculated using the WIA. The rates for most of the spectrum now satisfy $\alpha_L/\Omega \ll (2\pi)^{-1}$, but there still exists a substantial region for which $\alpha_L/\Omega > 1$, mostly in modes with high frequencies. This agrees with our general expectations; on the one hand we know from experience that the internal wave field appears to be weakly nonlinear, particularly for low frequencies. However we also know that all rates could not be small, otherwise the WIA would have been successful in the first place. We explicitly compare the WIA decay rates with SC rates in Figure (7) where we plot $\log(\alpha_L/a_L(WI))$. For small- m and small- Ω both calculations give roughly the same results. For moderate to large m we see substantially reduced rates for G . Notice though that in the small- m , large- Ω region the SC calculation actually increases rates by a great deal.

The change in rates for the moderate to large m region is easy to understand if one remembers that McComas (1977) and McComas and Bretherton (1977) identified the induced diffusion mechanism to be the dominant energy

transfer mechanism for the WIA in this region. This means that energy diffuses from a given large-wavevector mode to a nearby mode in wavevector space by interacting with a third, small-wavevector mode. Since in the WIA all waves are assumed to have infinite lifetimes the rate of energy exchange is only governed by the coupling between the modes. However in the SC calculation the finite lifetime of each mode is built into the calculation. This, and the fact that the decay rates depend on the amplitudes of the interacting modes, implies that the rate at which energy is transferred between modes decreases as the waves decay. Therefore we would expect the lifetimes of the modes in this region to be increased.

We would now like to consider another measure of the consistency of the calculation. Recall that we have found a correspondence between e and a given by (2.16). Given the parameters (a, c) we can now compute the ratio $e/2a_I U(0)$. In truth the degree to which $e/2a_I U(0)=1$ depends both on how well GM 76 represents an equilibrium spectrum and on the reliability of the calculation. It is not possible under the current assumptions to differentiate between these two effects, so it is best to talk about the degree of equilibrium of GM 76 under a given approximation. (See also the discussion at the end of section 2) For example, Figure (8) shows a plot of $e/2a_I U(0)$ for the WIA. We see that for large frequencies the spectrum appears to be in a high degree of equilibrium, while the lower frequencies are out of equilibrium by 50% or more. We compare this with Figure (9) which shows $\log(e/2a_I U(0))$ for the SC calculation. We note immediately that the previous discussion concerning the accuracy of α_L is relevant here, since a factor of 2 in the high- m region can greatly affect the equilibrium balance. The equilibrium nature of GM 76 may be difficult to determine for

large- m in this SC calculation. However in general we notice a similar kind of deviation of $e/2a_1 U(0)$ from 1, on the order of 50% or so, although where the WIA sees depressed values the SC method sees increased values. We take this to be a validation of the computational process, given the tentative nature of the GM spectra, which Garrett and Munk propose as "straw men" to be knocked down by the next generation of experiments.

Next, we have done another test calculation to determine how much the SC calculation might affect the off-resonant character of Σ_1 and Σ_2 . We have calculated the derivative of Σ_1^{-+} at $\omega=\Omega$. In Figure (10) we plot

$$\log \left| \Sigma_1'^{-+}(\Omega) \Omega / \Sigma_1^{-+}(\Omega) \right| \quad (4.4)$$

a measure of the rate of deviation of Σ_1^{-+} from its value at $\omega=\Omega$. If this quantity is much greater than 1 then the constant Σ_1 assumption breaks down. If we compare this with the same quantity calculated for the weak interaction case, shown in Figure (11), we see that the region where the relative deviation exceeds 1 is greatly diminished in the SC case. It should be noted in passing that the very restrictive nature of the weak interaction calculation (only allowing modes to interact if they satisfy the strict frequency resonance conditions) allows a new class of interaction to suddenly "turn on" at $\omega=2f$, and this accounts for the large derivatives around that frequency.

5. CONCLUSIONS

In this paper we have described a two pronged advance in the study of the time behavior of oceanic internal waves. First, we have applied a formalism which allows for a systematic study of internal wave interactions. Second we have performed the first self-consistent calculation of oceanic internal wave parameters. We will first discuss the importance of the formalism.

Since the work of Olbers (1974, 1976) and McComas (1977) showed that a naive perturbation theoretic approach to internal waves did not give reliable decay rates the possibility of improvement on their calculation has been discussed. Holloway (1977, 1978 1979) introduced a "frequency filter" correction. The problem is that by unsystematically introducing physically motivated corrections one leaves open the possibility that other corrections of comparable magnitude have been ignored. In some cases the combination of all of these possible corrections could actually yield results much different from those obtained from keeping only one correction. We believe that the new diagonal term, $c(k)$, in our calculation provides an example. The case where $a(k)=\text{finite}$, $c(k)=0$ for all k corresponds to the frequency filter correction. We found, however, that a nonzero $c(k)$ had both a qualitative and a quantitative effect on the results. First, a nonzero $c(k)$ allows for the possibility of an overdamped regime, and in fact we found that most of the modes in our calculation lie in the overdamped regime. Second,

the small rates caused by the subtraction of a_I and $|\Omega_R|$ in the overdamped regime give us a much different picture of the time behavior of internal waves than would be given simply by a_I , the decay rate dictated by the frequency filter alone. We therefore conclude that it may be difficult to include all significant corrections to the calculation from physical motivations alone, and that the formalism provides a systematic way to indicate the relevant effects.

The numerical results of this paper indicate the degree to which the earlier numerical calculations were unreliable. In fact we conclude that while that earlier decay rates were basically correct for the small wave-vector--small frequency regime, the decay rates for most of the spectrum were much too large. Our calculation indicates that, except for high frequencies, internal waves tend to decay relatively slowly. We point out that the fact that some region of the spectrum is still predicted to decay quickly is another validation of our procedure. Had all decay rates turned out to be small we would have been forced into the impossible conclusion that the weak interaction approximation was correct after all. Any future improvements on our calculation cannot change this fact. Further, this argument tends to imply that any future calculation cannot substantially reduce rates in the high frequency regime, since our numbers there are not much greater than the point where weak interaction theory becomes valid. This leads us to believe that our rates are fairly representative of reality, or at least represent a rough lower limit to it.

We close by considering how our SC calculation might be improved upon. The most logical next step is to study the effect of non-constant Γ_1 and Γ_2 on the results. This might be done by choosing functional forms for $\Gamma_1(\omega)$ that allow

for more than two poles in the Green's function. A continued fraction representation of Σ_1 has been tried in other contexts and would probably work well here. The problem with going to a larger number of poles is that it is no longer possible to find analytical expressions for the poles and so some insight might be lost. Further difficulties arise because a continued fraction representation allows for new poles to appear in G in successive iterations. Since it is not usually possible to keep all of these poles for computational reasons, one must determine a reasonable way to discard some of these poles. Nevertheless choosing some consistent scheme, like keeping the slowest decaying terms, would make such a calculation an attractive candidate for the next level of calculations.

This research is supported in part by contract number N00014-80-C-0840.

Figure Captions

1. Contour plot of $\log |c/(\Omega - a_R)|$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis. The contour interval is 0.4. The zero line corresponds to critical damping. All modes to the left of this line are underdamped; all lines to the right are overdamped. Dashed contours indicate lines with negative values. The horizontal axis is marked off in equivalent vertical wavelenths (in meters) and the vertical axis is marked off in units of the Coriolis frequency f .
2. Contour plot of $\log a_L(WI)/\Omega$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis. The contour interval is 0.2. Dashed contours indicate lines with negative values. The horizontal axis is marked off in equivalent vertical wavelenths (in meters) and the vertical axis is marked off in units of the Coriolis frequency.
3. Contour plot of $\log |\alpha_L/\alpha_g|$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis. The contour interval is 0.4.
4. Contour plot of $\Delta(a_L)$, the percentage deviation of the input value of a_L compared to the output value, vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis.
5. Contour plot of $\log \Delta(\alpha_L)$, the logarithm of the percentage deviation of the input value of α_L compared to the output value, vs. $\log(m)$ on the horizontal axis and $\log(m)$ on the vertical axis. The contour interval is 0.5.
6. Contour plot of $\log |\alpha_L/\Omega|$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis. The contour interval is 0.4.

7. Contour plot of $\log |\alpha_L/a_I(WI)|$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis. The contour interval is 0.5.
8. Contour plot of $e/(2a_I(WI) U(0))$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis. The contour interval is 0.4.
9. Contour plot of $\log(e/(2a_I U(0)))$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis. The countour interval is 0.4.
10. Contour plot of $\log |\Sigma_1^{i-+}(\Omega)\Omega/\Sigma_1^{i-+}(\Omega)|$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis, for the self-consistent calculation. The contour interval is 0.4.
11. Contour plot of $\log |\Sigma_1^{i-+}(\Omega)\Omega/\Sigma_1^{i-+}(\Omega)|$ vs. $\log(m)$ on the horizontal axis and $\log(\Omega)$ on the vertical axis, for the weak interactions case. The contour interval is 0.4.

REFERENCES

Carnevale, G.F. and Fredericksen, J.S. (1982)

"Viscosity renormalization based on direct interaction closure" LJI preprint.

DeWitt, R. J. (1982) "Self-consistent effective medium parameters for nonlinear random oceanic internal waves." PhD thesis Physics department University of Illinois, Urbana, Illinois

DeWitt, R. J. and Wright, Jon (1982) Self consistent effective medium theory of random internal waves. J. Fluid Mech. 115, 283-302.

Garrett, C. J. R. and Munk, W. H. (1972) Space-time scales of internal waves. Geophys. Fluid Dyn., 2, 225-264.

Garrett, C. J. R. and Munk, W. H. (1975) Space-time scales of internal waves: A progress report. J. Geophys. Res. 80, 291-297.

Holloway, G. and Hendershott, M. C. (1977) Stochastic closure for nonlinear Rossby waves. J. Fluid Mech. 82, 747-765.

Holloway, G. (1978) Order and disorder in nonlinear fluid motions. Topics in Nonlinear Dynamics-A Tribute to Sir Edward Bullard, American Institute of Physics, New York.

Holloway, G. (1979) On the spectral evolution of strongly interacting waves. Geophys. Astrophys. Fluid Dynamics 11, 271-287.

Kraichnan, R. H. (1959) The structure of isotropic turbulence at high Reynolds numbers. Fluid Mech. 5 497-543.

- McComas, C. H. (1977) Equilibrium mechanisms within the oceanic internal wave field. J. Phys. Oceanogr. 7, 836.
- McComas, C. H. and Bretherton, F. P. (1977) Resonant interactions of oceanic internal waves. J. Geophys. Res. 82, 1397.
- Olbers, D. J. (1974) On the energy balance of small-scale internal waves in the deep sea. Hamburg. Geophys. Einzelschr., no. 24, G. M. L. Wittenborn Sohnes, Hamburg.
- Olbers, D. J. (1976) Nonlinear energy transfer and the energy balance of the internal wave field in the deep ocean. J. Fluid Mech. 74, 375-399.
- Pomphrey, N., Meiss, J.D., and Watson, K. M. (1980) Description of nonlinear internal wave interactions using Langevin methods. J. Geophys. Res. 85 1085.

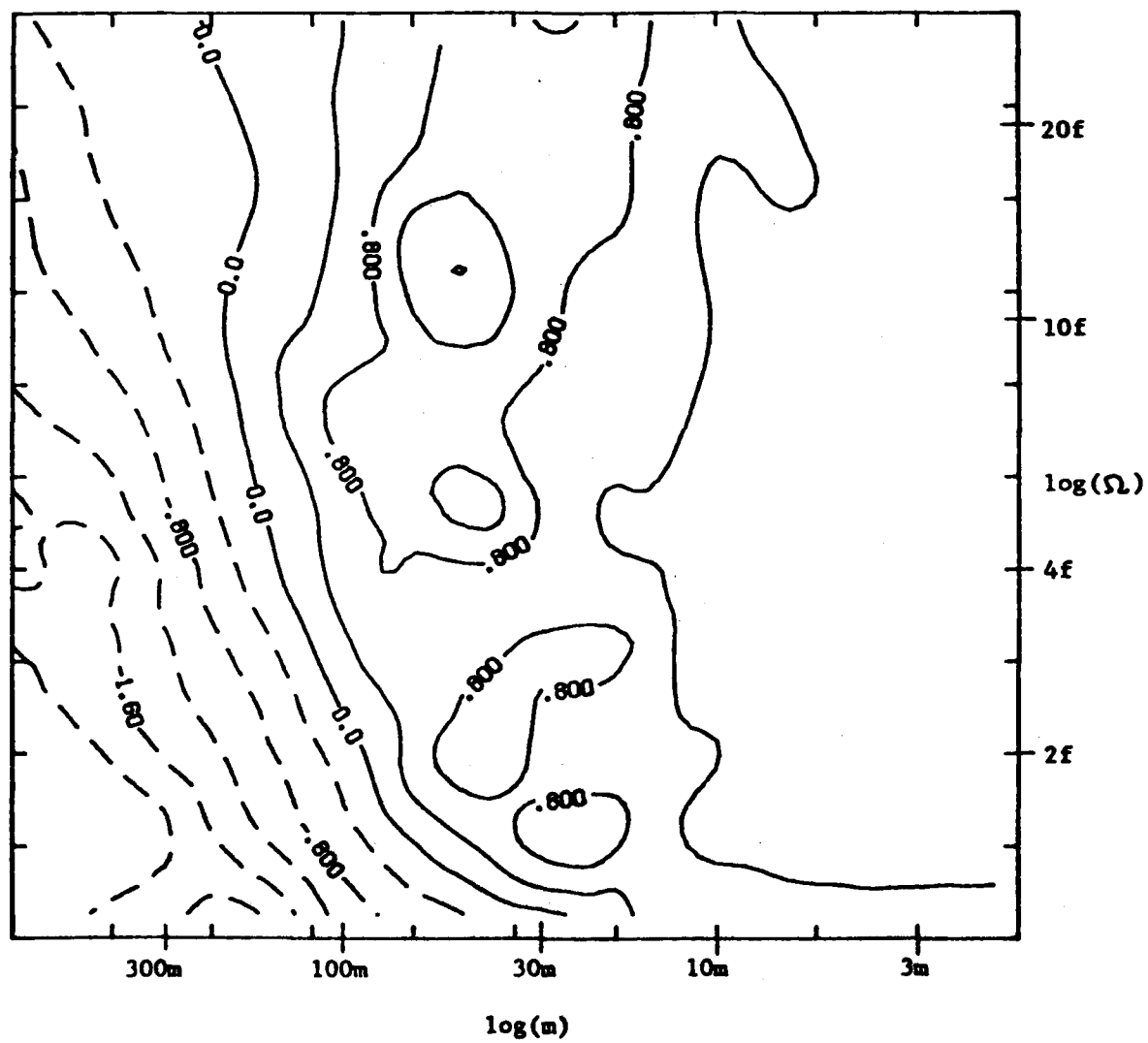


Fig. 1

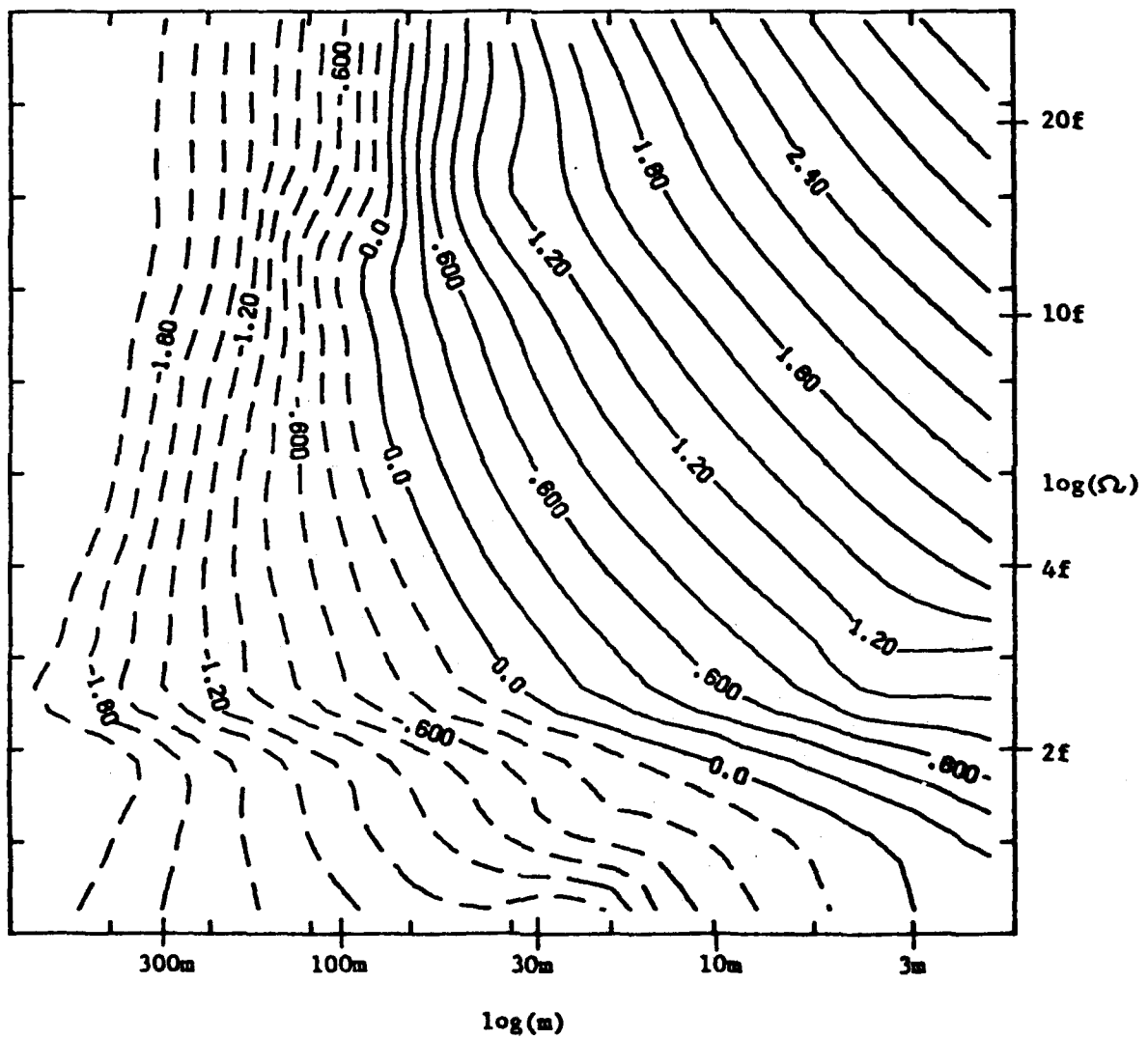


Fig. 2

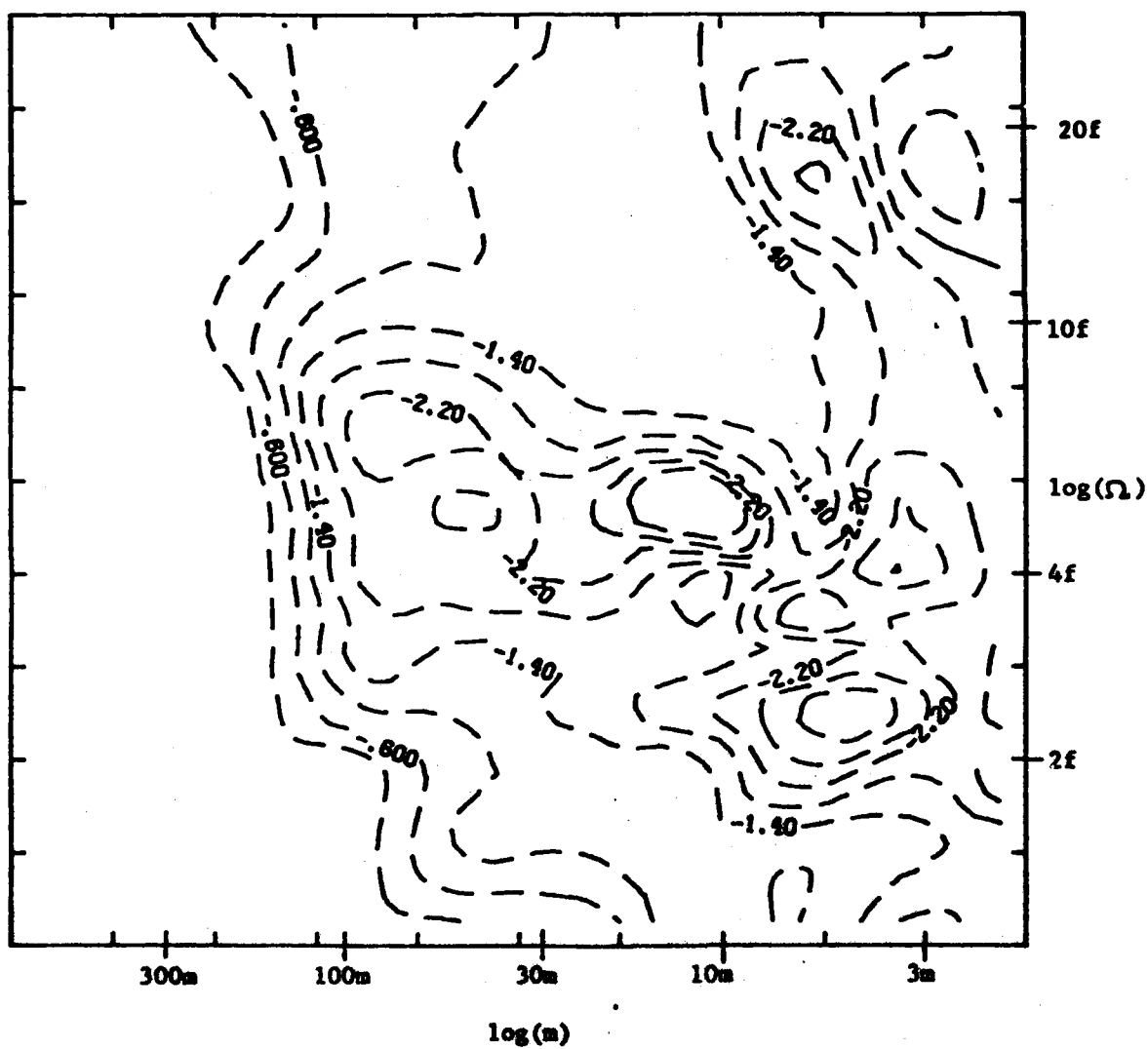


Fig. 3

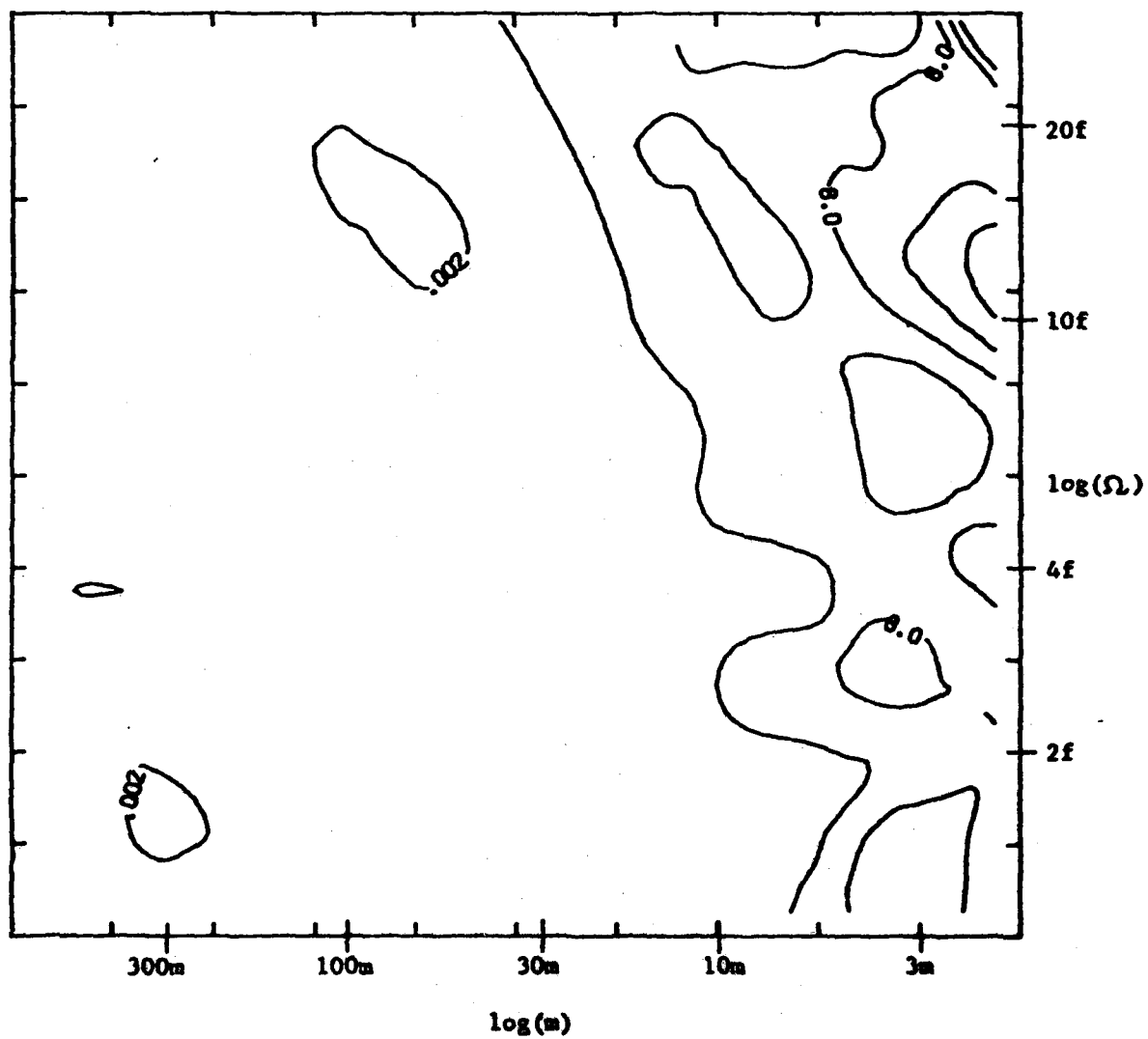


Fig. 4

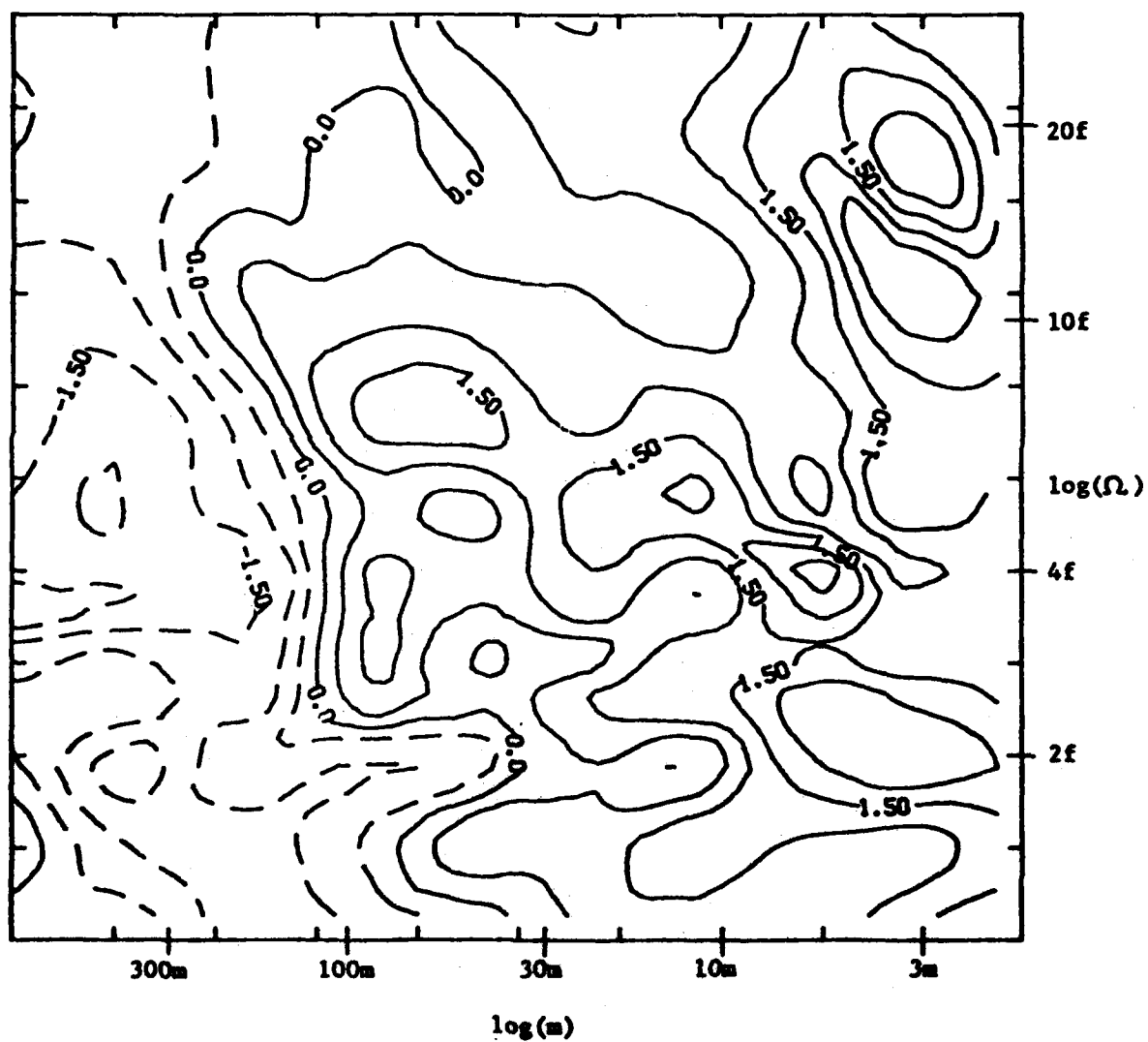


Fig. 5

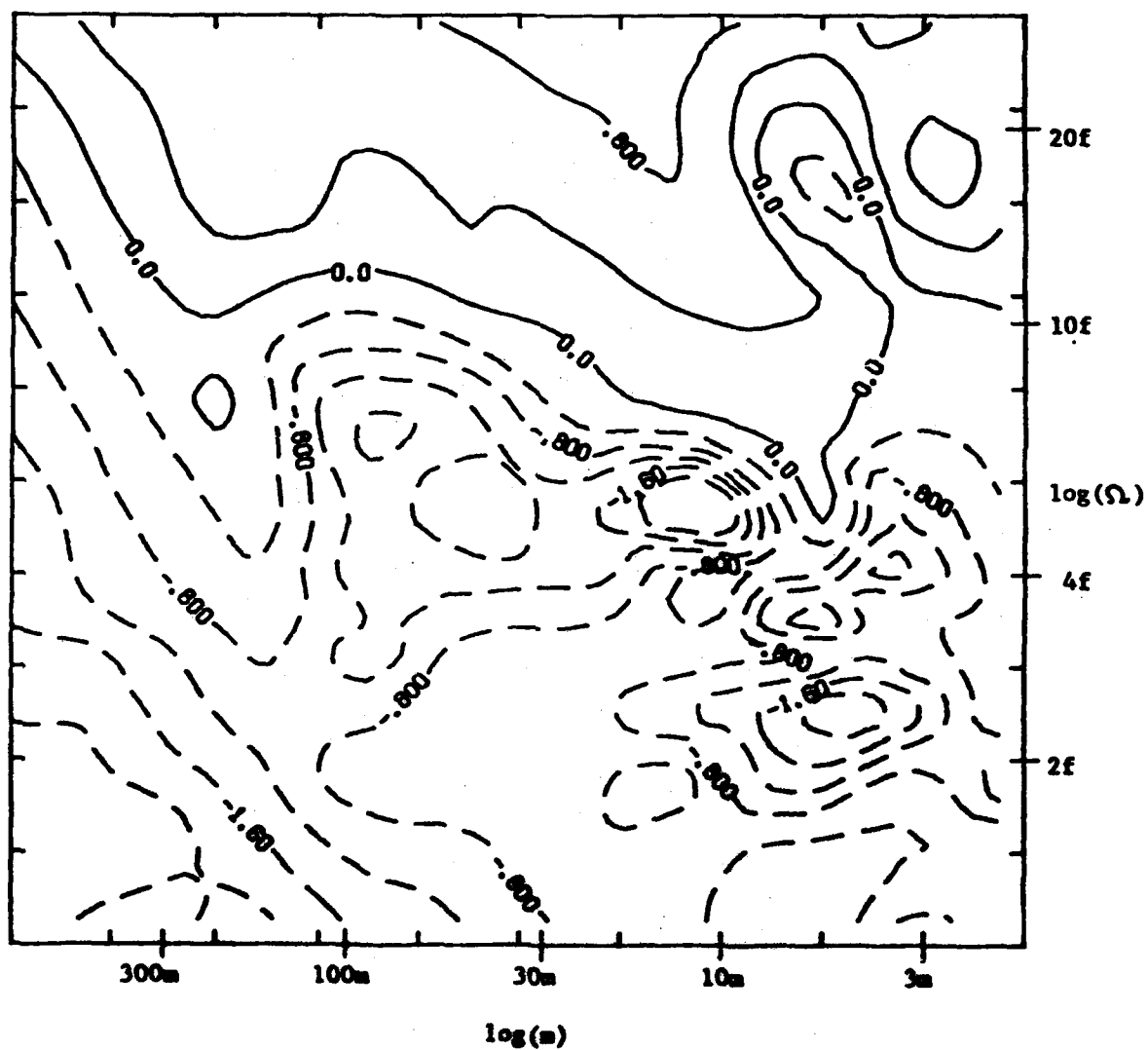


Fig. 6

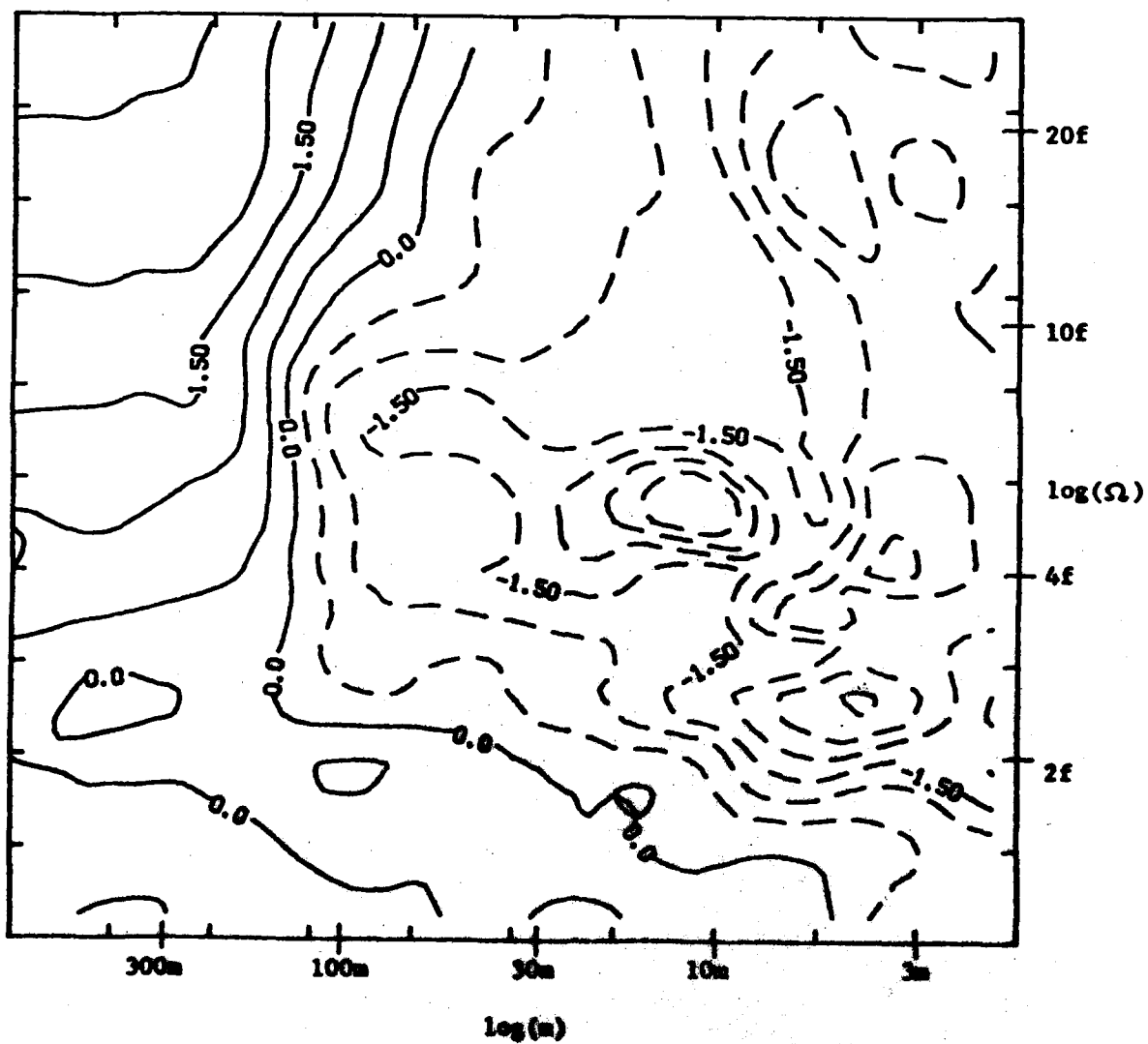


Fig. 7

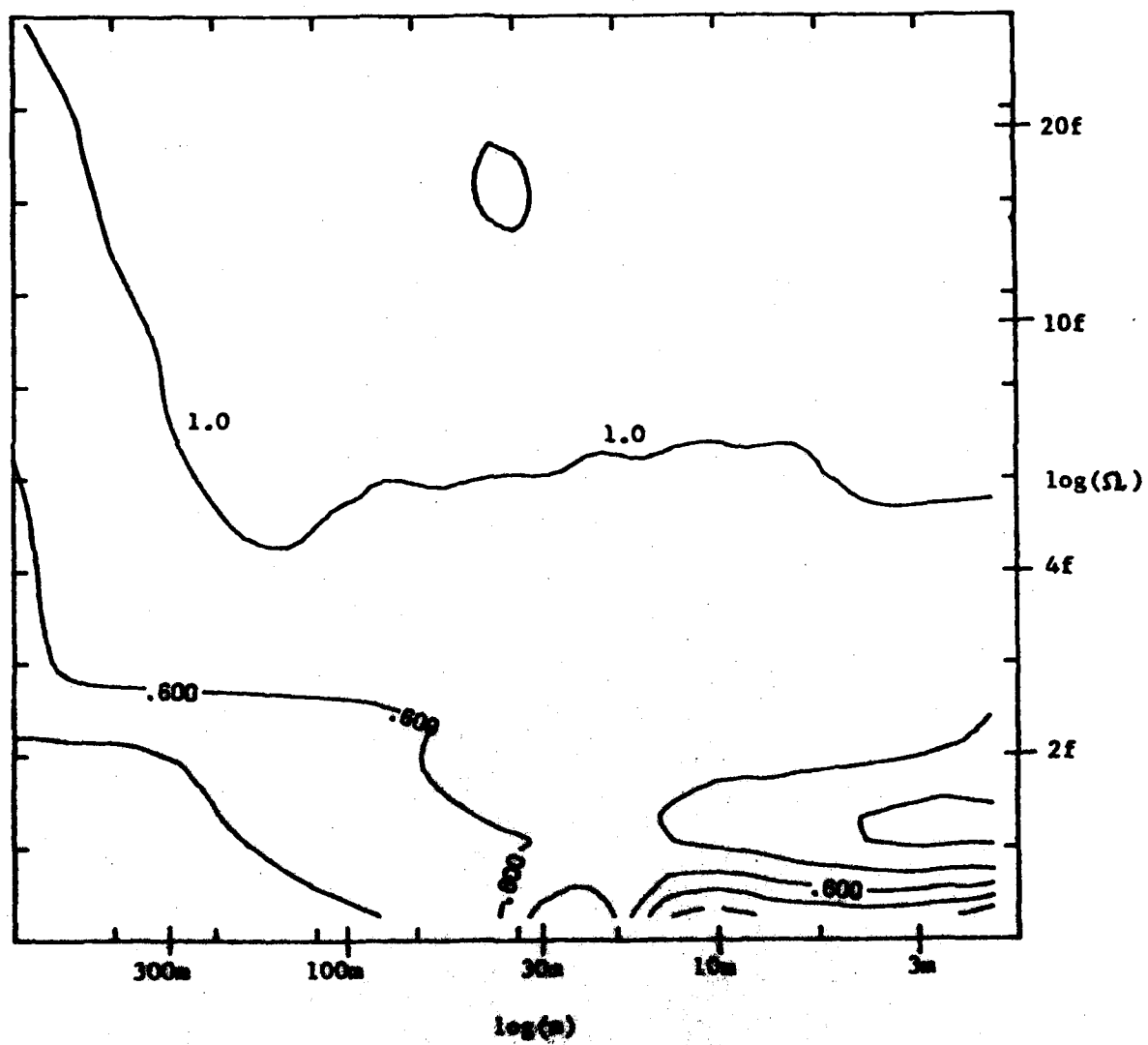


Fig 8

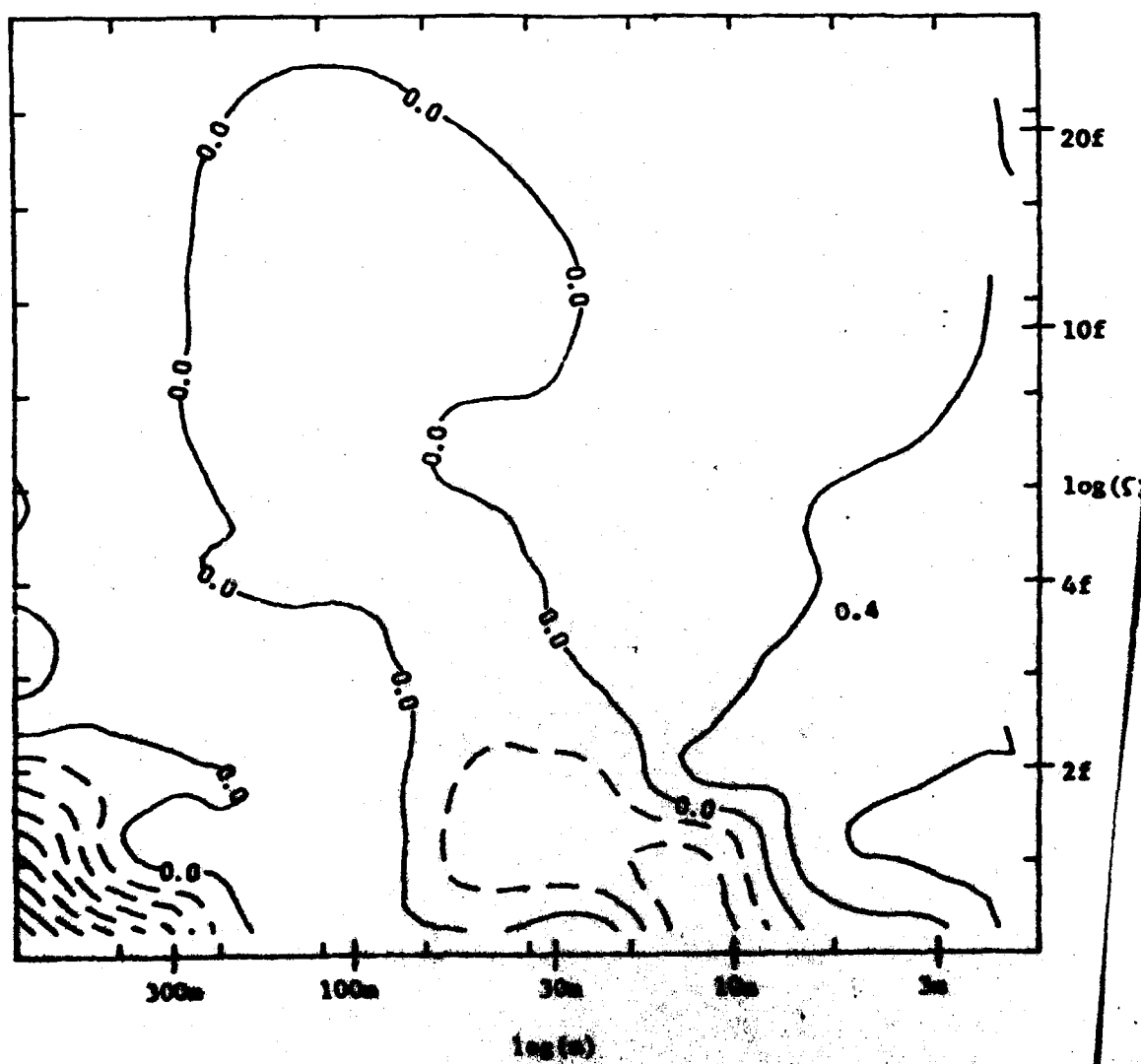


Fig 9

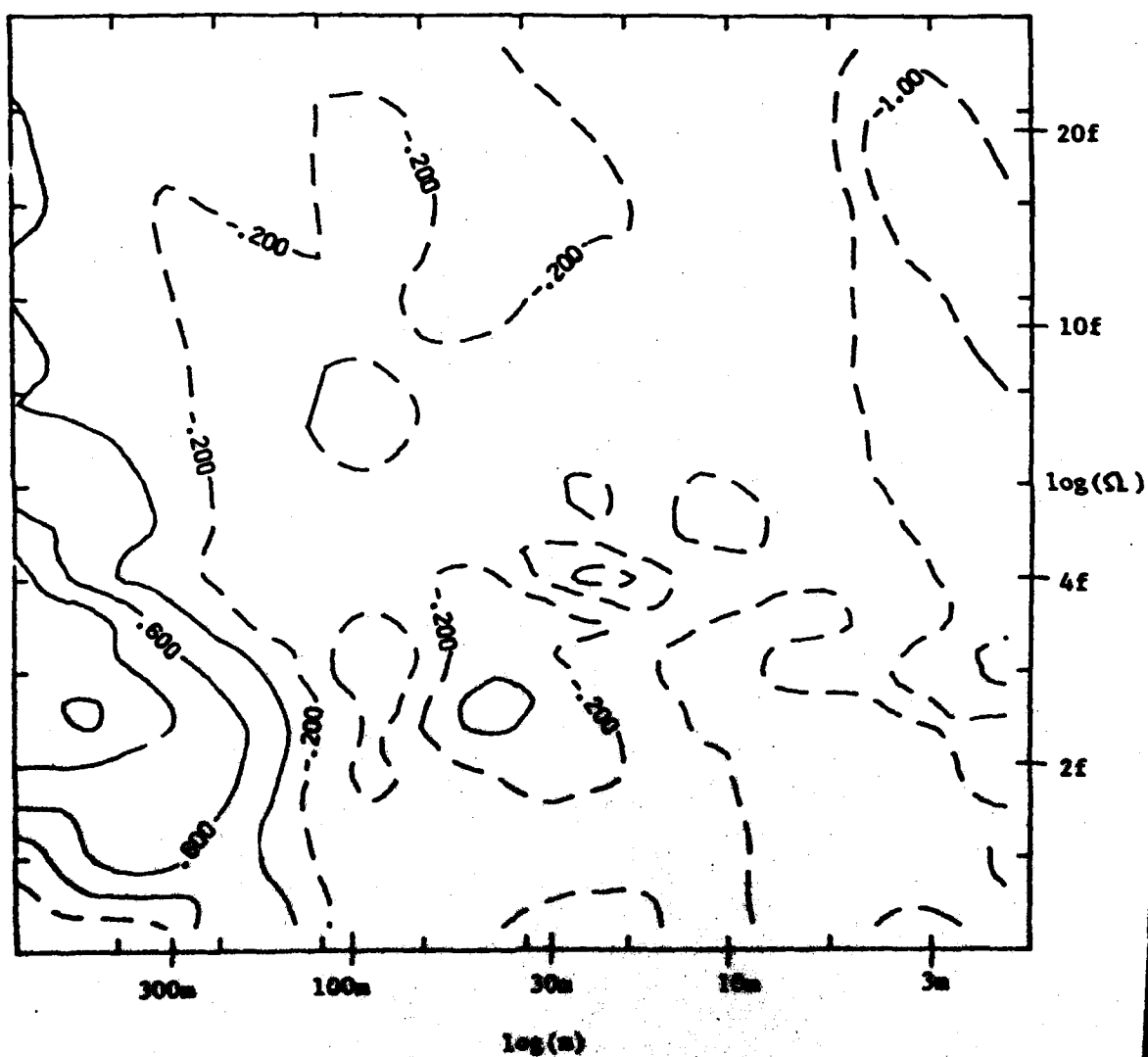


Fig. 10

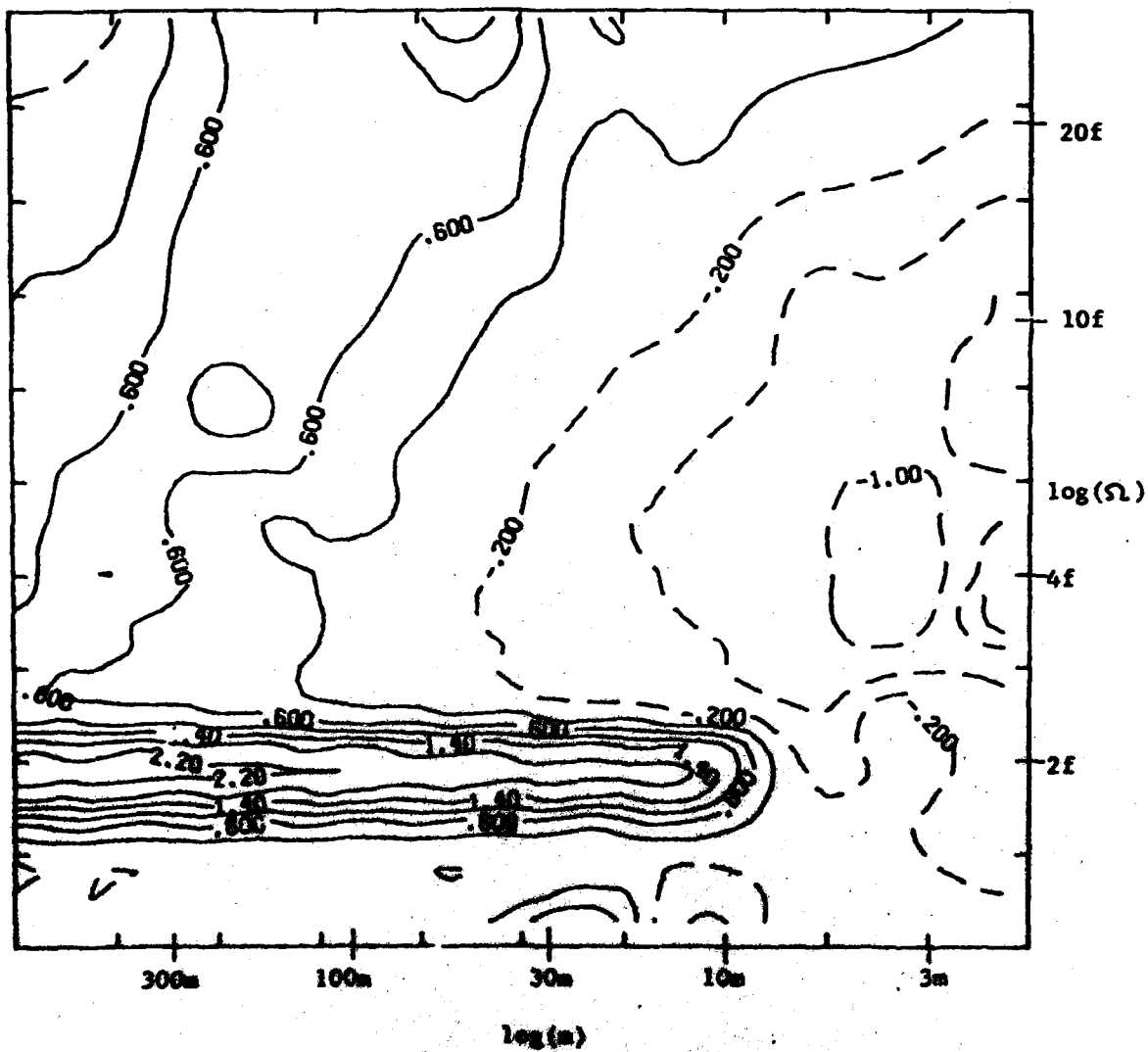


Fig. 11

ILL-(TH)-81-32
May, 1981

TRICRITICAL POINTS AND BIFURCATIONS IN A QUARTIC MAP

Shau-Jin Chang, Michael Wortis, Jon A. Wright

**Department of Physics
University of Illinois at Urbana-Champaign
1110 West Green Street
Urbana, IL 61801**

**To be published in the Proceedings of CNLS Conference "Nonlinear
Problems: Present and Future"**

**PACS Index: 02.50.+s
05.40.+j
05.70.Jk**

Typed by Mary Ostendorf

TRICRITICAL POINTS AND BIFURCATIONS IN A QUARTIC MAP

Shau-Jin Chang, Michael Wortis, Jon A. Wright

Department of Physics
University of Illinois at Urbana-Champaign
1110 West Green Street
Urbana, IL 61801

We have studied the 1-dimensional iteration map associated with the even quartic polynomial $x_{n+1} = 1 + a x_n^2 + b x_n^4$. This map allows a smooth transition from a single hump to a double hump. Bifurcations and higher-order transitions occur as we vary the parameters a, b . In addition to the usual universal bifurcation behavior discovered by Feigenbaum, we find a new universality class of bifurcations which is associated with a tricritical point. Tricritical points serve as natural boundaries to Feigenbaum critical lines. For the quartic map, the tricritical points which are the end-points of the original Feigenbaum line are $(a, b) = (0, -1.59490)$ and $(-2.81403, 1.40701)$. Associated with each tricritical point, there are two unstable directions as well as two independent exponents. The exponents are $\delta_T^{(1)} = 7.2851$ and $\delta_T^{(2)} = 2.8571$. At the tricritical point, we can introduce a universal function $f_T^*(x)$ which obeys

$$a_T f_T^*(f_T^*(x/a_T)) = f_T^*(x)$$

with the scale factor $a_T = -1.69031$. The quartic map has a special duality transformation $(a, b) \rightarrow (a', b')$, such that the two mappings are intrinsically related. The tricritical points which are dual to the above pair of tricritical points are located at $(-3.18980, 2.54371)$ and at $(.95561, -1.14981)$ and are joined by a line which is the dual of the original Feigenbaum line. There are an infinite number of different tricritical points which form at least a Cantor set.

I. INTRODUCTION

Recently, Feigenbaum has shown that the bifurcation sequence in a single hump iterative map $x_{n+1} = f(x_n)$ with a quadratic peak obeys a universal behavior.^{1,2} One can understand this behavior from a renormalization-group point of view. At the limiting point of a bifurcation sequence, the 2^n iteration of the map with appropriate scaling approaches a universal function $f_T^*(x)$. In the neighborhood of this universal function, it appears that there is only one relevant direction, along which the eigenvalue $\delta (= 4.6692)$ is larger than one. We have recently studied a more general map described by a quartic polynomial $x_{n+1} = f(x_n) = 1 + a x_n^2 + b x_n^4$. An important property of the quartic map is that $f(x)$ can describe both a single-hump and a double-hump map. By changing the parameters a, b continuously, one can induce a smooth transition from a one-hump map to a two-hump map and vice versa. It is easy to see that a double-hump map admits many different kinds of stable cycles which do not exist in the iterations of a single-hump map. An important concept in a double-hump map is that of doubly-stable cycles. These are cycles whose members pass through both the central peak and a side peak. In the vicinity of a doubly-stable cycle in the a, b parameter space, both peaks control the

stability of the same cycle. Away from doubly-stable cycles, however, the peaks control different cycles which are unrelated in x -space. Thus, the regions controlled by these different peaks become dynamically coupled at a doubly-stable cycle. This kind of dynamical coupling is a new phenomenon which cannot occur in a single-hump map. Its presence modifies the bifurcation processes. The limiting point determined by a bifurcation sequence of 2^N doubly-stable cycles is associated with a fixed point in function space with two relevant directions in analogy with Feigenbaum's discussion of the quadratic case. This limiting point also serves as the end-point of a critical line which describes the usual Feigenbaum behavior. This type of fixed point is known as a tricritical point in phase-transition language.³ The tricritical point discovered here has two universal exponents $\delta_T^{(1)} = 7.2851$, $\delta_T^{(2)} = 2.8571$. At the fixed point, the map is self-similar near the peaks after 2^N iterations. Just as in the Feigenbaum's case, it is possible to introduce a universal function $f_T^*(x)$ at a tricritical point which obeys the same functional equation

$$\alpha_T f_T^*(f_T^*(x/\alpha_T)) = f_T^*(x)$$

but with a different scale factor $\alpha_T = 1.69031$. Our $f_T^*(x)$ can be expressed as a power series in x^4 . We have discovered an identity, $\delta_T^{(2)} = \alpha_T^2$, relating one of the exponents to the scale factor. This identity is a member of a whole class of identities which exist at the fixed point of an iterative map.

II. QUARTIC MAPS

(A) Single vs multiple-hump maps

We study the one-dimensional iterative maps generated by an even quartic polynomial

$$x_{n+1} = f(x_n) \equiv 1 + a x_n^2 + b x_n^4. \quad (2.1)$$

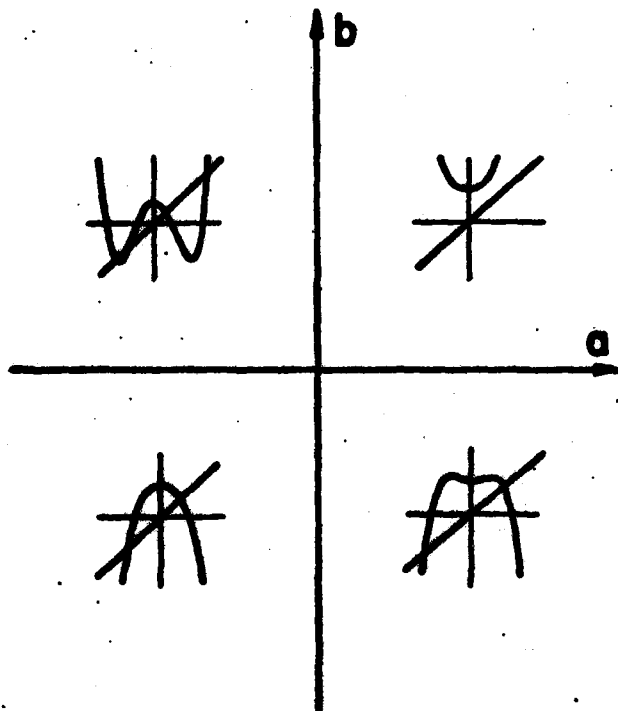


Figure 1

Iterative maps generated by the even quartic polynomial (2.1). The map in each quadrant corresponds to a $f(x)$ with (a, b) in that quadrant.

Depending on the values of a and b , $f(x)$ can have either a single hump or a double hump. See Fig. 1. For region defined by $a > 0$, $b > 0$, $f(x)$ can at most give rise to a stable fixed point. For $a < 0$, $b < 0$, the function $f(x)$ has a single quadratic hump, and Feigenbaum universality holds. In one parameter single-peaked maps, the limit point of infinitely bifurcated 2^N cycles is a single point. In two parameters single-peaked maps that point becomes a line that we refer to as the Feigenbaum line. For the regions $a < 0$, $b > 0$ and $a > 0$, $b < 0$, $f(x)$ has either two peaks and one valley or vice versa. In the following, we shall refer to the peak or valley at $x = 0$ as the central peak, and the peaks or valleys at $x \neq 0$ as the side peaks. The existence of both central and side peaks implies that $f(x)$ may develop independent iteration regions as indicated in regions I and II in Fig. 2.

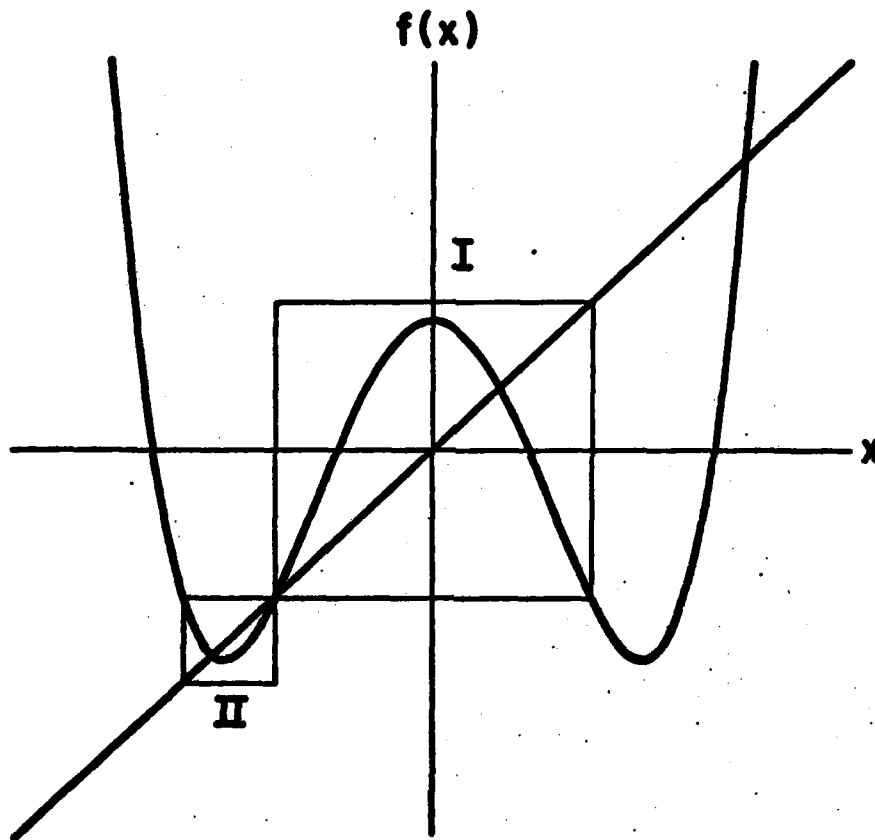


Figure 2
A map for $a < 0$, $b > 0$. Note that regions I and II are independent iteration regions.

In region I, $f(x)$ describes a single-hump map with a quadratic peak at $x = 0$. We shall refer to region I as the region controlled by the central peak. In region II, $f(x)$ describes an inverted single-hump map with the peak at $x = \pm 1/4b$. We shall refer to II as the region controlled by the side peak. Since the two side peaks have equal height, they map into the same x ($= -1/4b$). Hence, they cannot both be members of an N -cycle. Thus, we have at most one most stable region controlled by the side peaks. The existence of two independent regions may also appear in the case $a > 0$, $b < 0$.

(B) Duality transformation

For the study of cycles and their sequences, we only need to consider the region controlled by the central peak. The cycles associated with the side

peaks can be obtained from those with the central peak by a duality transformation described below.

Consider the quartic map (2.1). We can rewrite it as

$$x_{n+1} = 1 + a x_n^2 + b x_n^4, \quad b \neq 0$$

$$= 1 - \frac{a^2}{4b} + b \left(\frac{a}{2b} + x_n^2 \right)^2. \quad (2.2)$$

Thus, we can factor the quartic map as the product of two quadratic maps:

$$\xi_n = \frac{a}{2b} + x_n^2, \quad (2.3)$$

$$x_{n+1} = 1 - \frac{a^2}{4b} + b \xi_n^2. \quad (2.4)$$

We can now visualize the original mapping as

$$x_1 \rightarrow \xi_1 \rightarrow x_2 \rightarrow \xi_2 \rightarrow x_3 \rightarrow \dots \quad (2.5)$$

It is easy to see that if the x -mapping has a stable N -cycle, so does the ξ -mapping and vice versa. The mapping from ξ_n to ξ_{n+1} is simply

$$\xi_{n+1} = \frac{a}{2b} + \left(1 - \frac{a^2}{4b} + b \xi_n^2 \right)^2. \quad (2.6)$$

After a rescaling, we can put the ξ -mapping (2.6) into the standard form

$$\xi_{n+1} = 1 + a' \xi_n^2 + b' \xi_n^4 \quad (2.7)$$

with

$$a' = \frac{(4b-a^2)[8ab + (4b-a^2)^2]}{32b^2} \quad (2.8)$$

and

$$b' = \frac{[8ab + (4b-a^2)^2]^3}{(8b)^4}. \quad (2.9)$$

It is easy to verify that the side peaks of x are mapped into the central peak of ξ , and the side peaks of ξ are mapped into the central peak of x . The duality relation is reciprocal: The duality transformation of (a', b') is the original (a, b) .

(C) Most-stable cycles and doubly-stable cycles

For a quartic map, the most-stable N -cycles (x_1, x_2, \dots, x_N) are described by

$$f(x_i) = x_{i+1}, \quad x_{N+1} = x_1, \quad i = 1, 2, \dots, N \quad (2.10)$$

and

$$\frac{d}{dx} [f^N(x)]_{x_i} = 0. \quad (2.11)$$

An N -cycle is most stable, if either the central peak ($x = 0$) or one of the side peaks ($x = \pm \sqrt{-a/2b}$) is a member of the cycle x_i . The polynomials associated with a stable N -cycle give rise to a line in the a - b plane. In Fig. 3, the solid lines represent the loci of most-stable 2-cycles. Functions whose parameters satisfy $a + b + 1 = 0$ have most-stable 2-cycles associated with the central peak. The curved lines are the loci of points of most-stable 2-cycle associated with the side peaks. It is easy to see that, with the exception of the origin, the most stable N -cycle trajectories associated with the central peak never intersect among themselves. Neither do the most stable trajectories associated with the side peaks. However, the trajectories associated with the central peak do intersect those associated with the side peaks. Actually, they intersect in two completely different ways. At inter-

sections A and A' in Fig. 3, the most stable 2-cycles pass through both the central peak and one of the side peaks. We call this kind of cycle a doubly-stable cycle. Two trajectories which intersect at a doubly-stable cycle are "dynamically coupled" at the intersection. At intersections B and B', the 2-cycles are controlled by different peaks which are not related: some regions of x are attracted to one cycle and some to the other. We refer to this kind of intersection as "dynamically independent". Note that doubly stable points A and A' are dual to each other. The concept of dynamical coupling is invariant under duality transformation.

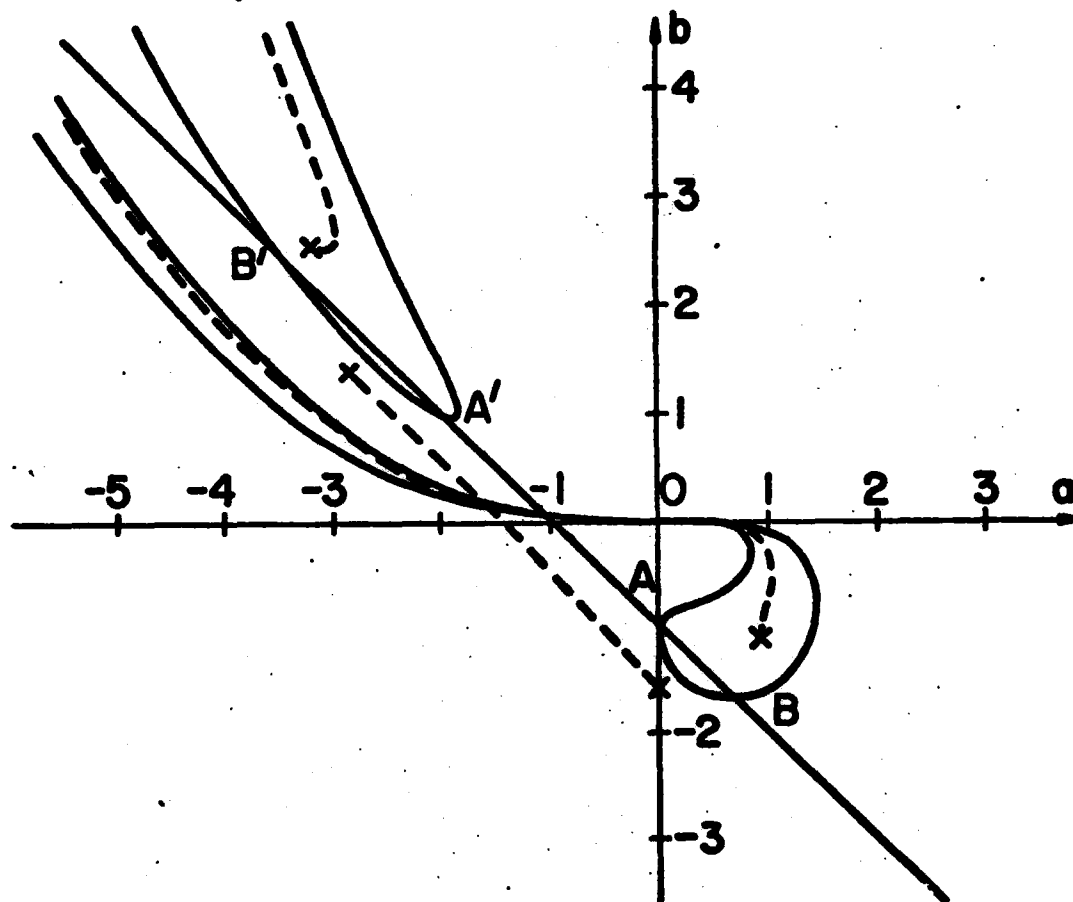


Figure 3

The solid lines represent the most-stable 2 cycles, the crosses denote tricritical points, and the dashed lines describe the Feigenbaum lines.

III. TRICRITICAL POINTS

(A) Bifurcation along b -axis

If we vary b (negative), keeping a small and negative, we encounter the usual bifurcation sequence, studied by Feigenbaum. If we vary b (negative), keeping a small and positive, we find that the map goes through three different most stable 2-cycles before bifurcating into a 4-cycle. These 2-cycles are controlled by the three different peaks as shown in Fig. 1. These phenomena persist no matter how small the value of a . The same phenomena repeat themselves at higher 2ⁿ cycles. Thus, there are qualitative differences between bifurcations on the two different sides of the b -axis. When $a = 0$ and $b < 0$, the curve $f(x)$ is single-humped with a quartic maximum. We find a sequence of bifurcations which always occur at the doubly-stable 2ⁿ cycles:

2^N	b_N
2	-1
4	-1.50393
8	-1.58225
16	-1.59316
32	-1.59466
...	...

The 2^N cycles have a limiting point at $b_\infty = -1.59490$, in close analogy to the quadratic map studied by Feigenbaum. However, because our map has a quartic (rather than quadratic) maximum, it is to be expected that there will be different critical exponents,⁴ as we now discuss. This limiting point has several interesting properties which we shall discuss below.

(B) Critical exponents and the universal function

There are two relevant eigen-directions at the critical point $(0, -1.59490)$. Along a relevant eigen-direction and for large N , we have

$$b_N = b_\infty + \frac{\text{const}}{\delta_T^N} \quad \text{with } |\delta_T| > 1 \quad (3.1)$$

where δ_T is the critical eigenvalue along this eigen-direction. One eigen-direction is along the b -axis where $\delta_T^{(1)} = 7.2851$. The other eigen-direction is $(1, -1.2347)$ where $\delta_T^{(2)} = 2.8571$. Note that the Feigenbaum fixed point has only one relevant direction with $\delta_F = 4.6692$. In phase-transition language, a critical point with two relevant directions which also serves as the end-point of a critical line is known as a "tricritical" point. As we shall see, our fixed point is indeed a tricritical point. See Discussion Section below.

At the critical point $(0, -1.59490)$, the function is self-similar near $x = 0$ to within a proper rescaling. The scale factor here is $\alpha_T = -1.6903$. Just as in the Feigenbaum case, the tricritical point has its own universal function $f_T(x)$ obeying the same functional equation

$$\alpha_T f_T^{(2)}(x/\alpha_T) = f_T^*(x), \quad \alpha_T = -1.6903. \quad (3.2)$$

In the neighborhood of $x = 0$, $f_T^*(x)$ can be represented approximately by

$$f_T^*(x) = 1 - 1.83413 x^4 + 0.01301 x^8 + 0.31188 x^{12} - 0.062035 x^{16} + \dots \quad (3.3)$$

(C) An identity

The second exponent $\delta_T^{(2)}$ and the scale factor α_T are related by

$$\delta_T^{(2)} = \alpha_T^2. \quad (3.4)$$

This relation is exact, as can be shown as follows: Let $h(x)$ be the eigenvector belonging to $\delta_T^{(2)}$. If

$$f(x) = f_T^*(x) + \epsilon h(x), \quad 0 < \epsilon \ll 1, \quad (3.5)$$

then iteration under

$$g(x) \equiv \frac{1}{f(1)} f^{(2)}(xf(1)) \quad (3.6)$$

leads by definition to

$$g(x) = f_T^*(x) + \delta_T^{(2)} \epsilon h(x) + O(\epsilon^2). \quad (3.7)$$

Since $h(0) = 0$ (because $f(0) = 1$), we may take

$$h(x) = x^2 + c_4 x^4 + c_6 x^6 + \dots \quad (3.8)$$

Direct iteration of (3.6) gives $(\alpha_T = 1/f_T^*(1))$

$$\begin{aligned} \alpha_T^{-1} \delta_T^{(2)} h(x) = & -h(1) f_T^*(x) + h(f_T^*(\frac{x}{\alpha_T})) \\ & + [h(1) \times f_T^{*'}(\frac{x}{\alpha_T}) + h(\frac{x}{\alpha_T})] f_T^{*'}(f_T^*(\frac{x}{\alpha_T})) \end{aligned} \quad (3.9)$$

which is a functional equation for $h(x)$. It is easy to verify that the right-hand side vanishes at $x = 0$. Only the last term contributes to $O(x^2)$ and by equating both sides of (3.9) one finds

$$\delta_T^{(2)} = f_T^{*'}(1)/\alpha_T. \quad (3.10)$$

One evaluates $f_T^{*'}(1)$ by differentiating (3.2) with respect to x and then setting $x = 0$. The result is

$$f_T^{*'}(1) = \alpha_T^3, \quad (3.11)$$

which gives (3.4). The other eigenvalues belonging to directions out of the $x^4, x^8, x^{12} \dots$ subspace are irrelevant. Eq. (3.4) is an example of a general class of identities which we discuss elsewhere.

(D) The tricritical point and doubly-stable cycles

Our present tricritical point $(0, -1.59490)$ is the bifurcation limit of a sequence of doubly-stable 2^n cycles. This turns out to be a general property of a bifurcation tricritical point. We have used this property to determine many other tricritical points which do not lie on the b -axis.

IV. DISCUSSION

In the $a < 0, b < 0$ region, we always have a single-hump function $f(x)$. If we increase the parameters $-a, -b$, we encounter an infinite sequence of bifurcations, as described by Feigenbaum. The limiting points of these infinite bifurcations now lie on a line in the a - b plane which we refer to as a Feigenbaum line. The tricritical point $(a, b) = (0, -1.59490)$ serves as a natural boundary to the original Feigenbaum line. When we extend the original Feigenbaum line in the other direction, we find that it terminates at another tricritical point, located at $(-2.81402, 1.40701)$. One can check easily that this tricritical point is also the limit of a sequence of doubly-stable 2^n cycles and that it has the same critical exponents $\delta_T^{(1)}$ and $\delta_T^{(2)}$.

Under the duality transformation, a tricritical point maps into a tricritical point and a Feigenbaum critical line maps into a Feigenbaum line. The dual images of the previous tricritical points are located at $(-3.18980, 2.54371)$ and at $(0.95561, -1.14981)$. We have plotted the dual transformed Feigenbaum line in Fig. 3.

Since bifurcation occurs after each stable cycle, and since there are an infinite number of different stable cycles, there are an infinite number of Feigenbaum lines in the a - b plane. Hence, there must also be an infinite number of tricritical points in the a - b plane. Indeed, we are able to show that between the critical points $(0, -1.59490)$ and $(.95561, -1.14981)$, there are an infinite number of tricritical points, forming a Cantor set. We shall discuss the geometrical meaning of the Cantor set elsewhere.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant Nos. NSF PHY79-00272, NSF DMR78-21069, and by the Office of Naval Research under Contract No. N00014-80-C-0840.

REFERENCES

. 8

1. Feigenbaum, M. J., J. Stat. Phys. 19, 25 (1978); 21, 669 (1979), Phys. Lett. 74A, 375 (1979).
2. For reviews on iterative maps, see e.g. May, R. H., Nature 261, 459 (1976); Collet, P. and Eckmann J. P., Iterated Maps on the Interval as Dynamical Systems (Birkhäuser Co., Boston 1980).
3. The term "tricritical point" was originally introduced by R. B. Griffiths to describe the two-fluid critical mixing point: Griffiths, R. B., Phys. Rev. Lett. 24, 715 (1970). See also, Riedel, E. K., Phys. Rev. Lett. 28, 675 (1972); Griffiths, R. B., Phys. Rev. B7, 545 (1973).
4. Derrida, B., Gervois, A. and Pomeau, Y., J. Phys. A12, 269 (1979).

Transitions and distribution functions for chaotic systems

Shan-Jin Chang and Jim Wright

Physics Department, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

(Received 22 September 1980)

We study chaotic systems generated by deterministic or probabilistic mappings. We introduce the density function which is an eigenfunction of a probability-preserving kernel K . We are able to show that all eigenvalues of K have magnitude less than or equal to 1 and that the only magnitude-one eigenvalues are the N th roots of unity. We have also calculated the corresponding eigenfunctions associated with these magnitude-one eigenvalues. These eigenfunctions can be expanded in terms of N positive functions having disjoint support. We then concentrate on a one-dimensional system, and study the behavior and mechanism for various chaotic transitions. We find that the mechanism associated with the 2 to 1 (or more generally, $2N$ to N) transition is different from those associated with other chaotic transitions. We then determine the conditions for these transitions, and express them in a universal form. We confirm the Feigenbaum-Raschke scaling in the large 2^n to 2^{n-1} chaotic-transition region, and determine the prefactor at these transitions. In addition, we establish a simple relation between the Lyapunov exponent and the folding of the distribution functions. We have also studied the chaotic regions of this system numerically.

I. INTRODUCTION

The problem of chaotic behavior in deterministic nonlinear systems has received considerable attention in recent years.¹⁻⁵ In particular one-dimensional maps have been extensively studied as there is increasing evidence that under certain circumstances complicated systems have an underlying one-dimensional structure. For a recent review of one-dimensional maps see the article by May.¹ The features of interest are the sequences of bifurcations leading to 2^n cycles of the basic periods and the subsequent chaotic behavior which has a similar bifurcation structure. All sequences of bifurcations occur as some parameter is changed. Smaller and smaller changes of the parameter lead to limit cycles with larger n . Finally an accumulation point is reached and for values of the parameter beyond that point the process is inverted, except that the behavior is now characterized by cycling between regions instead of isolated points. These sequences of bifurcations are characterized by a remarkable scaling property discovered by Feigenbaum.² He discusses the scaling for the isolated fixed points. Feigenbaum and Raschke³ point out that the same scaling behavior applies in the chaotic region as well. The original purpose of this investigation was to investigate what changes in the theory result if the mapping is allowed to be slightly probabilistic. Some information on this point is discussed by Crutchfield and Feigenbaum.⁴ The theory we developed turned out to provide us with some new insights into the deterministic one-dimensional map, and consequently part of this paper discusses these developments.

It seems appropriate to study some of the properties of stochastic maps, and eventually to be

more precise about the sources and size of the stochasticity as related to some experimental configuration. In this paper we set up a formalism for discussing the equilibrium-density distribution and the approach to equilibrium. In the next section we will describe the formalism and demonstrate certain general properties of a wide class of models. Then in the third section we will restrict ourselves to a deterministic one-dimensional map and we will exhibit a number of simple features of such maps. We discuss the transitions between 2^n cycles and 2^{n-1} cycles and then the new behavior that occurs in the three to one transition. In the large 2^n to 2^{n-1} chaotic-transition region, we confirm the Feigenbaum-Raschke scaling³ and are able to determine the prefactor associated with these chaotic transitions. Some other simple scaling properties are discussed as well. A simple expression for the divergence of neighboring points (Lyapunov exponents) is given in Sec. IV. The numerical method used is discussed briefly in Appendix A.

II. GENERAL THEORY

We consider the mass of the unit interval onto itself by a probability-preserving mapping,

$$x \rightarrow y = f(x), \quad (2.1)$$

where f is a probability-preserving mapping of the unit interval onto itself. We assume that f is a piecewise linear mapping, and that f is invertible. We assume that f is a piecewise linear mapping, and that f is invertible. We assume that f is a piecewise linear mapping, and that f is invertible.

$$K(x, y) \geq 0. \quad (2.2)$$

For a technical reason, basically aimed at ruling out general unitary kernels associated with area-preserving mappings, we also need the operator K to have an additional property known to mathematicians as quasicompact. Equation (2.3) immediately gives

$$\int \psi(x) dx = \int \phi(x) dx. \quad (2.4)$$

A surprising amount of information can be obtained from the conditions (2.3) and (2.5). In particular we will show that the equidistributed distribution is characterized by N ($N=1, 2, 3, \dots$) disjoint regions all having the same area. The operator K maps cyclically the function characterizing one of these regions into the function for another region. These statements (with some qualifications) are independent of the dimensionality of the system. Note that the operator in (2.1) can correspond to a set of differential equations with no time-dependent parameters or to a set with periodic parameters. The mapping would then be for a fixed time interval. For our numerical examples we will use

$$x_{n+1} = \lambda x_n(1-x_n) \quad (2.5)$$

which gives for K ,

$$K(x, y) = \delta[x - \lambda y(1-y)]. \quad (2.6)$$

One way to introduce probabilistic behavior is to assume that at each iteration λ is chosen according to a probability distribution $P(\lambda)$. Then we would have for K ,

$$K(x, y) = \int P(\lambda) \delta[x - \lambda y(1-y)] d\lambda.$$

A simple version of this is to take K to be the sum of two terms

$$K(x, y) = p\delta[x - \lambda_1 y(1-y)] + (1-p)\delta[x - \lambda_2 y(1-y)]. \quad (2.7)$$

Note that this mechanism for adding probabilistic behavior is different from the one studied in Ref. 4. To study the solutions to Eq. (2.1), we will imagine finding the eigenfunctions and eigenvalues of K . We suppose that K is reasonably well behaved, although operators such as the one in Eq. (2.6), are acceptable provided we allow δ -function distributions. The stable periodic system corresponds to δ -function eigenfunctions. The eigenfunction problem for K involves finding both left and right eigenfunctions

$$\int K(x, y) \psi(y) dy = \eta \psi(x) \quad (2.8)$$

and

$$\int L(x) K(x, y) dx = \eta L(y). \quad (2.9)$$

By integrating Eq. (2.8) and using Eq. (2.2) we see that if $\eta \neq 1$,

$$\int K(x, y) dx = 0.$$

Next take the absolute value of Eq. (2.8) to obtain

$$\int K(x, y) |\psi(y)| dy \geq |\eta| |\psi(x)|. \quad (2.10)$$

We now have the following results:

Lemma 1: All eigenvalues η of K satisfy

$$|\eta| \leq 1.$$

This result follows immediately from integrating Eq. (2.10) over x .

Lemma 2: If $R(x)$ is a right eigenfunction of K with eigenvalue $|\eta| = 1$, then $\phi_R(x) = |R(x)|$ is an eigenfunction with eigenvalue 1.

Proof. From Eq. (2.10) and the fact that $|\eta| = 1$ we have

$$D(x) = \int K(x, y) \phi_R(y) dy - \phi_R(x) \geq 0.$$

Integrating $D(x)$ over x and using Eq. (2.2), we have

$$\int D(x) dx = 0.$$

This implies $D(x) = 0$, and consequently

$$K\phi_R = \phi_R.$$

Lemma 3: If η is an eigenvalue of K with eigenfunction $|\eta| = 1$, and if there is no degeneracy for the eigenvalue 1, we have

(a) There is no integer N such that $\eta^N = 1$. This says that all eigenvalues of absolute value 1 are not roots of unity.

(b) There are N nonnegative functions ψ_i with eigenvalue η^i such that

$$K\psi_0 = \psi_0, K\psi_1 = \eta\psi_0, \dots, K\psi_{N-1} = \eta^{N-1}\psi_0.$$

(c) If η is a root of unity, then there are N nonnegative functions ψ_i with eigenvalue η^i such that

$$K\psi_0 = \psi_0, K\psi_1 = \eta\psi_0, \dots, K\psi_{N-1} = \eta^{N-1}\psi_0.$$

(d) If η is a root of unity, then there are N nonnegative functions ψ_i with eigenvalue η^i such that

$$K\psi_0 = \psi_0, K\psi_1 = \eta\psi_0, \dots, K\psi_{N-1} = \eta^{N-1}\psi_0.$$

(e) There are N eigenfunctions of K having eigenvalues with absolute value 1. Define $\alpha = e^{2\pi i/N}$.

$$x_{n+1} = f(x_n)$$

with

$$K(x, y) = \delta[x - f(y)].$$

Since the mapping is from a point of zero width to another point of zero width, the iteration of a point may never return to itself. However, in the chaotic region a small neighborhood will return and cover the original neighborhood after a sufficient number of iterations. If we assume that the phase of the function $\phi(x)$ is piecewise continuous, then it is possible to establish Lemma 3 even with a δ -function K . We shall modify our definition of regions by assuming that there are small neighborhoods which lie entirely inside these regions. Finally we can show that if the eigenvalue equal to one is degenerate then there will be disjoint groups mapping onto themselves, but not onto members of other groups and there will be one such group for each eigenvalue equal to unity. For a general mathematical discussion of some of the above ideas, see Ref. 6.

We have computed some eigenvalues and eigenfunctions of K . An example of an eight cycle is shown in Fig. 1 for the K of Eq. (2.6). In later sections we will show more examples of the distribution function, paying particular attention to transition regions. An important feature of Fig. 1 is the narrowness of most of the regions (they all have equal area under the curve). In an experimental setting it would be difficult to distinguish a chaotic eight cycle from an eight cycle with δ -

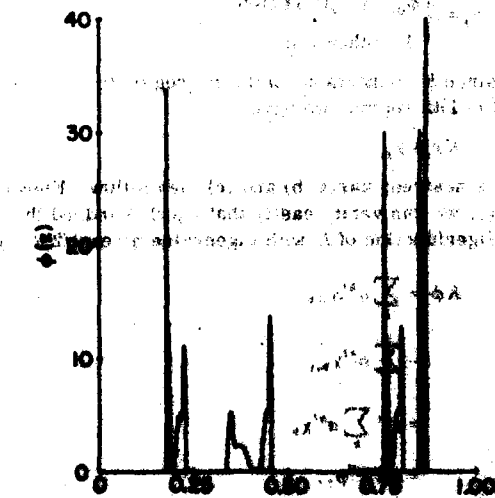


FIG. 1. A small random term has been added to the K of Fig. 1 to smooth out the distribution. The eigenvalues are $\lambda_1 = 0.98$, $\lambda_2 = 3.571$, and $\lambda_3 = 3.5$. The distribution is symmetric about $x = 0.5$.

function supports. They will both have an underlying periodicity which will give similar results in experiments such as the one Feigenbaum² discusses.

The effect of adding some randomness to the system is illustrated in Figs. 2 and 3. In Fig. 2 we added a very small extra term to the K that gave Fig. 1. In the notation of Eq. (2.7) we chose $\rho = 0.98$, $\lambda_1 = 3.571$, and $\lambda_2 = 3.5$, the latter value corresponding to periodic cycles. The small amount of random behavior changed the eight-cycle regions of Fig. 1 into the two-cycle regions of Fig. 2. Chaos also results if two periodic maps are added together. We choose two λ 's, each of them corresponding to a periodic two cycle and added the maps according to Eq. (2.7). The resulting chaotic behavior is shown in Fig. 3.

We wish to question here the general behavior of the eigenfunctions. If K is bounded and piecewise continuous, the eigenfunction ϕ will at least be piecewise continuous. For a deterministic system $x_{n+1} = f(x_n)$, $K(x, y)$ is given by

$$K(x, y) = \delta[x - f(y)].$$

The eigenfunction is either a sum of δ functions or an ill-behaved function ϕ . In the latter case, if the maximum of $f(x)$ is quadratic, ϕ will in general consist of an infinite number of square-root singularities of the form $c/(1-x)^{1/2}$. The function ϕ is integrable, but not continuous, nor bounded. However, the phase of ϕ is usually continuous. In our figures we have cut off the highest of the peaks that were the numerical approximation

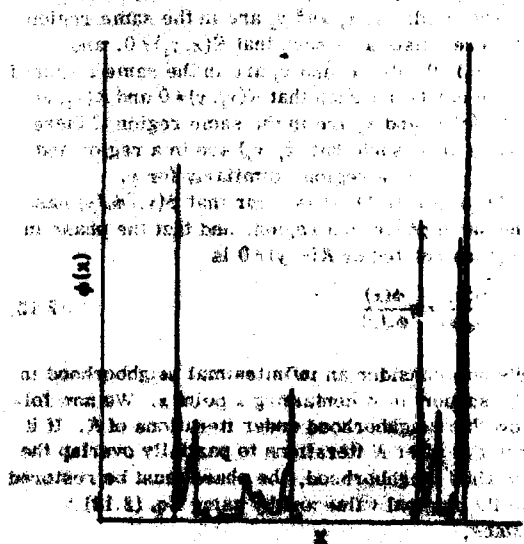


FIG. 2. A small random term has been added to the K of Fig. 1 to smooth out the distribution. The eigenvalues are $\lambda_1 = 0.98$, $\lambda_2 = 3.571$, and $\lambda_3 = 3.5$. The distribution is symmetric about $x = 0.5$.

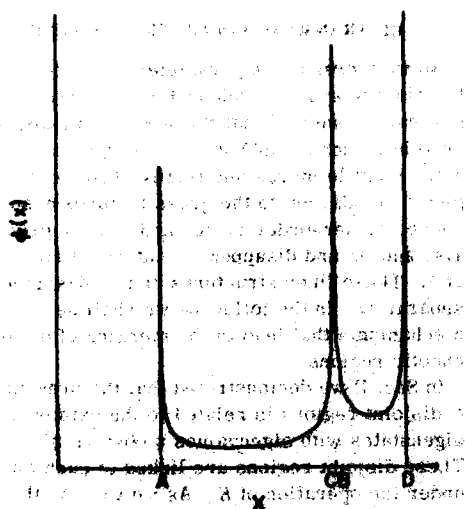


FIG. 8a. Just past the transition from 2 to 1, $\lambda = 2.579$.

originally separated regions. As we shall see, both mechanisms are realized for Eq. (3.5). We present an example of such mechanism:

1. 2^m to 2^{m-1} merging. We first consider the $2-1$ merging. We discussed our result trivially to cover the more general 2^m-2^{m-1} case. At the transition from two chaotic regions to a single chaotic region, we considered the first mechanism discussed there. Outside the position of the maximum of $f(x)$ by P , the peak position is mapped in subsequent iterations on the right of A .

f : $P \rightarrow$ point D — point A — point C

— point B — interior of CD .

The points are shown in Fig. 4. It is important that P lies in the interior of AB and that point B folds into the interior of AP during the mapping. We can easily see that region AB is mapped into CD and vice versa. In the gap BC there are two special points E, F (see Fig. 4). F is an unstable fixed point lying inside the gap BC and E is the other point such that $f(E) = C$. During an iteration, points inside the line segment BE are mapped into the interior of CD and are removed from the gap BC . Under repeated iterations all points in the gap are gradually moved into the segment BE and then removed from the gap. It is easy to see that points to the left of point A and to the right of point B are also mapped into regions AB and CD after repeated iterations. As we increase the coupling constant λ , points B, C , and F all converge and the two regions become joined.

at CB and we have only one region as shown in Fig. 5. The requirement that $f^2(P)$ is a fixed point determines the transition value of λ . Thus we have

$$f^{(2)}(P) = f^4(P). \quad (3.1)$$

In terms of $F(x) = f^2(x)$ we can reexpress this as

$$F^2(P) = F^4(P) = F(P). \quad (3.2)$$

As we shall see this equation applies for both transition mechanisms.

We can generalize our method to study the 2^m to 2^{m-1} chaotic region merging processes. If we consider $f^m(x)$ with $N = 2^{m-1}$, then the 2^m-2^{m-1} transition in f is a $2-1$ transition in f^m . Defining

$$F(x) = f^m(x) = f^{(m)}(x), \quad (3.3)$$

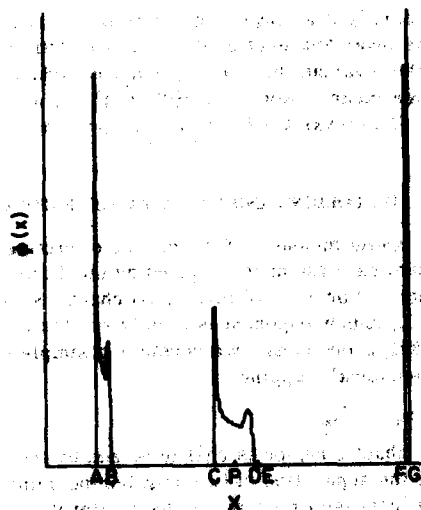
we find that the condition for 2^m to 2^{m-1} transition can also be written as Eq. (3.2) with $F(x)$ given by (3.3). The set of λ_m for 2^m-2^{m-1} transition based on the simple iteration equation $x_{n+1} = \lambda x_n(1-x_n)$ is listed in Table I.

2. 3 to 1 chaotic transition. Next, we consider the $3-1$ transition. Refer to Figs. 6 and 9 and define three regions I is (A, B) , II is (C, D) , and III is (E, G) . It is easy to see that the mechanism described above cannot work. In Fig. 9 we find that the three regions are mapped into each other according to $I \rightarrow II \rightarrow III \rightarrow I$. Let us assume (illegally) that the first mechanism does occur. As these three regions start to overlap in a manner similar to that shown in Figs. 7 and 8, their intersections $I \cap II$, $II \cap III$, $III \cap I$ will map into each other to form a three cycle. Since regions I and III do not intersect, this scheme is not possible. Hence, only the second mechanism described in Figs. 6-9, is allowed.

In a typical three-region chaotic map, Fig. 6, the chaotic regions and the peak position $x = P$ are mapped according to

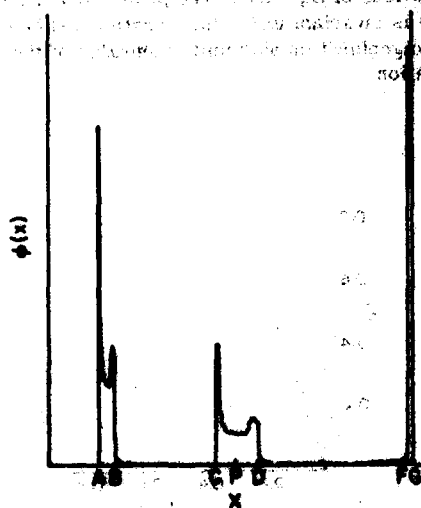
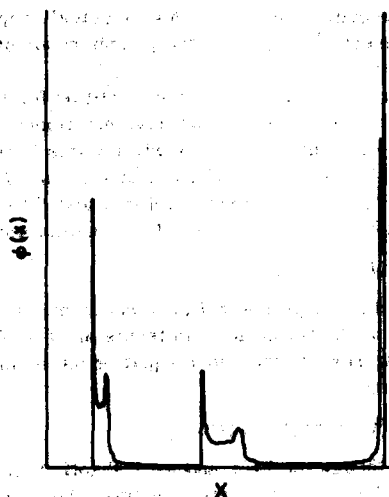
$$f: P \rightarrow G \rightarrow A \rightarrow C \rightarrow F \rightarrow E \rightarrow D \rightarrow H,$$

where H is inside FG . For a singly peaked $f(x)$, folding can only occur around the peak $x = P$. In

FIG. 6. Chaotic three cycle $\lambda = 3.858$.

terms of regions, we find that region I is mapped onto region II, region II after folding is mapped onto region III, and region III is mapped back onto region I.

There are two mechanisms for removing points from the gaps BC and DE . The first mechanism is through folding and mapping. Point C is folded onto point E (i.e., E also maps onto F ; see Fig. 9) and the segment DE is thus mapped into the interior of FG . The line segment DE serves as a "drain" to the gap BC through which the points in the gap may escape. The remaining points in BC are mapped into EF . Upon a further iteration, EF

FIG. 7. An asymptotic chaotic state $\lambda = 3.857$.FIG. 8. Past transition to single region $\lambda = 3.858$.

maps into BF , which contains both the gaps BC , DE , and region II. Of course, all points which map into II are removed from the gaps forever. This serves as an alternative "sink" for removing points from the gaps.

As λ changes, we may encounter the situation where D and E coincide. Then, the drain DE is sealed, but the other drain still works. Hence, we still have two gaps and three separate regions. However, as λ increases further, the locations of D and E are interchanged. The sink now becomes a source, and points in region II can map into the gaps via the opening at ED (see Fig. 9). Of course, we still have the other drain to remove points from the gaps. The combined effect gives rise to a finite distribution in the gaps which represents the equilibrium distribution between the source and the sink just as in the case of a flooded basement. Since the seepage is controlled by the size of the gate, we expect that the value of ϕ inside the gaps increases continuously from zero with the increase-

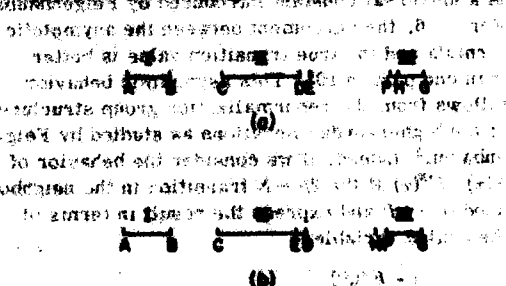


FIG. 9. (a) The ordering of the points discussed in the text for a chaotic three cycle. (b) The same order after the transition to a single cycle.

ing seepage. Figures 6-8 give actual graphical representations of ϕ in the neighborhood of a 3-1 transition.

Now, let us determine the condition for the 3-1 chaotic transition. As we have mentioned earlier, the 3-1 transition occurs when points D and E coincide. Since D is $f^4(P)$, and since point E maps into $F=f^4(P)$, the condition for D and E to coincide is for $f(D)=f(E)=F$. This then implies

$$f^4(P)=f^4(P). \quad (3.4)$$

Note that the points B, D, F form an unstable three cycle at the transition. In terms of $F(x)=f^4(x)$, we can rewrite the above equation as the universal equation

$$F^4(P)=F^2(P)+F(P). \quad (3.5)$$

Even though the 2-1 and 3-1 chaotic transitions arise from different mechanisms, they obey the same universal equation. The only difference is in the definition of F : F is f^2 for the 2-1 transition, and is f^4 for the 3-1 transition. We can generalize the above result to an arbitrary transition. For $f=\lambda x(1-x)$, the 3-1 chaotic transition occurs at

$$\lambda = 3.856\ 800\ 652\ 477\ 765, \dots$$

3. Universal behavior. We now discuss the universality of the F^2 to F^3 transitions. In Table I, we see that as $\lambda \rightarrow \infty$, λ_n approaches the infinite two-cycle bifurcation point $\lambda_\infty = 3.856\ 800\ 671 \dots$ from above. In fact, λ_n satisfies the asymptotic behavior

$$\lambda_n = \lambda_\infty + \frac{A}{n^2} \quad (3.6)$$

with

$$A = 0.001\ 484 \dots$$

and

$$A = 4.499\ 201\ 990 \dots$$

is a universal constant introduced by Feigenbaum.² For $n > 6$, the agreement between the asymptotic formula and the true transition value is better than one part in 10^6 . This asymptotic behavior follows from the renormalization group structure of the higher-order iterations as studied by Feigenbaum.² Indeed, if we consider the behavior of $F(x)=f^4(x)$ at the $2N-N$ transition in the neighborhood of $x=P$ and express the result in terms of the scaled variable

$$y = \frac{x - F^4(P)}{F^4(P) - F^2(P)},$$

$$G(y) = \frac{F^4(x) - F^4(P)}{F^4(P) - F^2(P)} = \frac{F^4(F^2(P) + (F^4(P) - F^2(P))y) - F^4(P)}{F^4(P) - F^2(P)}$$

we arrive at an invariant function $G(y)$ at large N . This invariant function G is a linear combination of the invariant functions g and h introduced by Feigenbaum.² Note that $G(0)=G(1)=0$, and $G(1/2)=1$. The invariant function $G(y)$ is shown in Fig. 10.

IV. FOLDING AND THE LYAPUNOV EXPONENT

A useful measure of the degree of mixing of a dynamical system is provided by the Lyapunov exponent. For a one-dimensional chaotic system, the Lyapunov exponent is simply related to the folding phenomena. We consider a simple one-dimensional mapping

$$x_{n+1} = f(x_n). \quad (4.1)$$

The chaotic region is characterized by the fact that the separation of two neighboring points (x_n, y_n) will increase exponentially as we iterate (4.1) a large number of times. For an infinitesimal separation, we refer to the average logarithmic increment defined by

$$\mu = \lim_{n \rightarrow \infty} \frac{1}{n} \lim_{x_0 \rightarrow y_0} \ln \left| \frac{x_n - y_n}{x_0 - y_0} \right| = \lim_{n \rightarrow \infty} \frac{1}{n} \left(\sum_{i=0}^{n-1} \ln |f'(x_i)| \right) \quad (4.2)$$

as the Lyapunov exponent. The exponent μ is positive ($\mu > 0$) in the chaotic region. Let us start from a small neighborhood. After many iterations, the neighborhood will enlarge and eventually cover a compact space. Further iterations will map this compact space into itself. It is known that there exists a density function $\phi(x)$ such that $\phi(x)dx$ provides the probability measure for finding an arbitrary point inside the region dx after many iterations of Eq. (4.1). The probability function $\phi(x)$ is invariant under the iteration and hence is the eigenfunction with unit eigenvalue of the linear equation

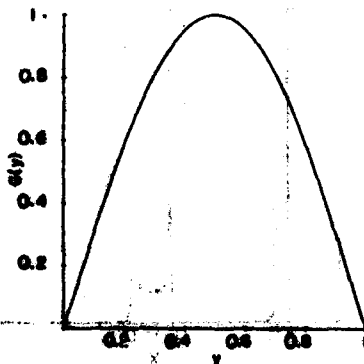


FIG. 10. The universal function $G(y)$.

$$\int K(x, y) \phi(y) dy = \phi(x), \quad (4.3)$$

where

$$K(x, y) = \delta[x - f(y)]. \quad (4.4)$$

We have learned from Lemma 3 of Sec. II that the support of ϕ may consist of N disjoint regions. The iteration of $K(x, y)$ [or (4.1)] will map points from one region to another, and bring the point back to the original region after N iterations. Thus, if we consider the iteration defined by

$$x_{n+1} = F(x_n), \quad (4.5)$$

with

$$F(x) = f^N(x) = f(f^{N-1}(x)), \quad (4.6)$$

we encounter iterations which map a region back onto itself. Since the Lyapunov exponent μ associated with f is related to the exponent μ_N associated with F via

$$\mu = \frac{1}{N} \mu_N, \quad (4.7)$$

we only need to consider F and μ_N involving simpler mappings. A typical distribution function is shown in Fig. 11. Let AB be a region which maps into itself under F . The end point B is mapped onto point A , and point A and an interior point D are both mapped onto E [$F(A) = F(D) = E$], and point P is mapped onto point B . This mapping may be achieved by folding PD on PA and then stretching BP to the original length AB (see Fig. 12). For more complicated mappings, it may be necessary to fold the line segment more than once. All the

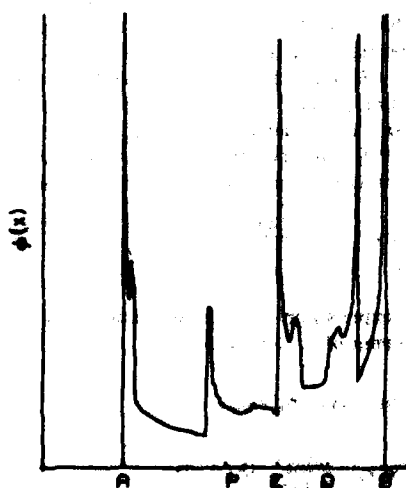


FIG. 11. Illustration of overlapping. Density function $\phi(x)$ for $\lambda = 2.75$.

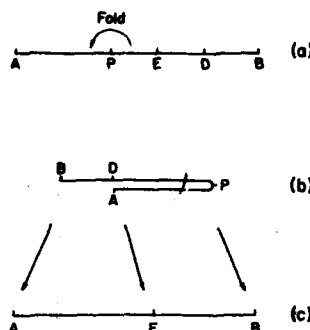


FIG. 12. Mapping involved in a chaotic one region: (a) Folding of AB about point P gives (b). Stretching BP completes the mapping shown in (c).

points in segment AE come from BD and the mapping is 1 to 1. We shall call this segment AE as nonoverlapping. The points in segment EB come from both AP and PD and the mapping is 2 to 1. We call this segment EB as overlapping.

We denote the density function ϕ within segments AP , PD , and DB as $\phi_{AP}(x)$, $\phi_{PD}(x)$, and $\phi_{DB}(x)$, respectively, obeying

$$\phi_{AP}(x) = \begin{cases} \phi(x), & \text{for } x \in (AP) \\ 0, & x \text{ outside } (AP) \end{cases} \quad (4.8)$$

etc. Then, we have

$$K^N \phi_{DB} = \phi_{AE}, \quad (4.9)$$

$$K^N \phi_{AP} = (\phi_1)_{EB}, \quad (4.10)$$

$$K^N \phi_{PD} = (\phi_2)_{EB}, \quad (4.11)$$

with

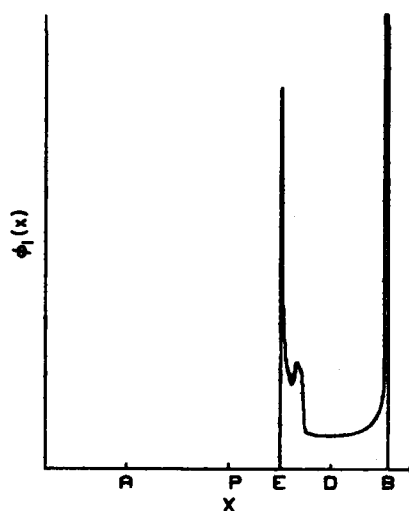
$$\phi_{EB} = (\phi_1)_{EB} + (\phi_2)_{EB}. \quad (4.12)$$

Equations (4.10) and (4.11) may be taken as the definitions of ϕ_1 and ϕ_2 . In the following, we shall express the Lyapunov exponent as a function of ϕ , ϕ_1 , and ϕ_2 . The ϕ_1 and ϕ_2 for the ϕ of Fig. 11 are shown in Figs. 13 and 14.

It is more convenient to denote the separation of two neighboring points as $\phi(x)dx$ rather than as dx . To understand the definition of m using ϕ , consider a particular point x_1 and its neighborhood dx_1 . Map it a sufficient number of times N such that it returns to its original neighborhood. If the new length is dx_1 , the magnification is

$$m = \left(\frac{dx_N}{dx_1} \right)^{1/N} = \left(\frac{\phi(x_1)dx_1}{\phi(x_N)dx_N} \right)^{1/N} = \left(\frac{\phi(x_1)dx_1}{\phi(x_2)dx_2} \frac{\phi(x_2)dx_2}{\phi(x_3)dx_3} \dots \frac{\phi(x_{N-1})dx_{N-1}}{\phi(x_N)dx_N} \right)^{1/N}.$$

The enlargement and reduction due to the extra factor $\phi(x)$ cancels and we have the same overall magnification factor. The advantage of introducing this modified magnification factor is that the magnification factor is always one in the nonoverlapping

FIG. 13. Density function ϕ_1 (see text).

ping region due to the conservation of the probability under the mapping $x = F(y)$, giving

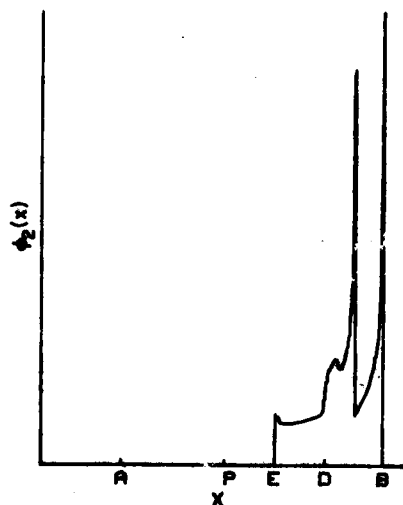
$$\phi(x)dx = \phi(y)dy \quad (4.13)$$

or

$$m = \frac{\phi(x)dx}{\phi(y)dy} = 1. \quad (4.14)$$

However, in the overlapping region, x can come from two or more different points, such as y_1 and y_2 ,

$$x = F(y_1) = F(y_2) \quad (4.15)$$

FIG. 14. Density function ϕ_2 (see text).

with

$$\phi_1(x)dx = \phi(y_1)dy_1, \quad (4.16)$$

$$\phi_2(x)dx = \phi(y_2)dy_2. \quad (4.17)$$

The magnification factors due to the mappings $y_1 \rightarrow x$ and $y_2 \rightarrow x$ are no longer one, and are given by

$$m(y_1) = \frac{\phi(x)dx}{\phi(y_1)dy_1} = \frac{\phi(x)}{\phi_1(x)}, \quad (4.18)$$

$$m(y_2) = \frac{\phi(x)dx}{\phi(y_2)dy_2} = \frac{\phi(x)}{\phi_2(x)}. \quad (4.19)$$

The geometrical average of the magnification factor m approaches

$$\bar{m} = \lim_{n \rightarrow \infty} \left(\prod_i m(y_i) \right)^{1/n},$$

and the Lyapunov exponent becomes

$$\mu_N = \ln \bar{m} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i \ln m(y_i). \quad (4.20)$$

To compute the limiting average magnification, we note that the weighting factor at any point y is proportional to its density $\phi(y)dy$. Thus, we can replace \sum_i in (4.20) by an integral over $\phi(y)dy$ and n by $\int dy \phi(y)$, giving

$$\ln \bar{m} = \frac{\int_{AB} dy \phi(y) \ln m(y)}{\int_{AB} dy \phi(y)}.$$

In the numerator of the above equation, we can separate the y integration into three regions: Region DB which maps into the nonoverlapping region (AE) of x gives no contribution since $\ln m = 0$. The other y 's are separated into y_1 and y_2 regions (AP and PD), and they both map into the overlapping region (EB) of x . The Lyapunov exponent becomes

$$\mu_N = \ln \bar{m} = \frac{\int dy_1 \phi(y_1) \ln m(y_1) + \int dy_2 \phi(y_2) \ln m(y_2)}{\int_{AB} dy \phi(y)}.$$

The expression has a nicer form when expressed in terms of x . Using

$$\phi(y_h)dy_h = \phi_h(x)dx,$$

$$m(y_h) = \frac{\phi(x)}{\phi_h(x)}, \quad (h = 1, 2)$$

we finally obtain

$$\mu_N = \ln \bar{n} = \frac{\int_{AB} dx \left(\phi_1(x) \ln \frac{\phi(x)}{\phi_1(x)} + \phi_2(x) \ln \frac{\phi(x)}{\phi_2(x)} \right)}{\int_{AB} dx \phi(x)}$$

$$= \frac{\int_{AB} dx \phi(x) \left(\frac{\phi_1(x)}{\phi(x)} \ln \frac{\phi(x)}{\phi_1(x)} + \frac{\phi_2(x)}{\phi(x)} \ln \frac{\phi(x)}{\phi_2(x)} \right)}{\int_{AB} dx \phi(x)} \quad (4.21)$$

The expression in the last large parenthesis of the numerator of (4.21) has the form of an entropy associated with the mixing of two independent statistical systems. We can generalize Eq. (4.21) trivially to include the folding of three or more regions.

For the folding of two regions, we introduce an overlapping factor r by

$$r(x) = \left(\frac{\phi_1(x)}{\phi(x)} \ln \frac{\phi(x)}{\phi_1(x)} + \frac{\phi_2(x)}{\phi(x)} \ln \frac{\phi(x)}{\phi_2(x)} \right) \frac{1}{\ln 2}. \quad (4.22)$$

It is easy to see that $r=0$ if either ϕ_1/ϕ or ϕ_2/ϕ is zero, and r has a maximum, $r_{\max}=1$, at $\phi_1/\phi = \phi_2/\phi = \frac{1}{2}$. In terms of r and ϕ , we have

$$\mu_N = \int_{AB} dx \phi(x) r(x) \ln 2 / \int_{AB} dx \phi(x). \quad (4.23)$$

To compute μ_N (or μ), we need detailed information about ϕ_1 , ϕ_2 , and ϕ as functions of x . However, we can set a rather good upper bound on μ_N if we know the integrals of ϕ_1 and ϕ_2 in the overlapping regions. Let

$$M_1 = \int_{AB} dx \phi_1(x), \quad M_2 = \int_{AB} dx \phi_2(x), \quad (4.24)$$

and

$$M = \int_{AB} dx \phi(x). \quad (4.25)$$

we then define a mean overlapping factor as

$$R = \frac{M_1}{M_1 + M_2} \ln \frac{M_1 + M_2}{M_1} + \frac{M_2}{M_1 + M_2} \ln \frac{M_1 + M_2}{M_2}. \quad (4.26)$$

One can show that

$$\mu_N \leq \frac{(M_1 + M_2)}{M} R \ln 2. \quad (4.27)$$

For the particular coupling leading to the distribution functions of Figs. 9-11, we obtain $\mu=0.36$ and the upper bound would give $\mu < 0.42$.

We have studied numerically the magnification factor and the Lyapunov exponent for the universal function G defined via Eq. (3.7) by sampling 300 000 iterations under G . The average magnification

m_G , the Lyapunov exponent μ_G , and the average overlapping factor R_G are

$$m_G = 1.9872,$$

$$\mu_G = \ln m_G = 0.6867,$$

and

$$R_G = 0.9907,$$

respectively. The large magnification ($m \approx 2$) and overlapping factor ($R \approx 1$) are not surprising. Numerically, the universal function $G(y)$ is very close to the quadratic function $f(y) = 4y(1-y)$. The eigenfunction corresponding to the latter mapping is $\phi = 1/[x(1-x)]^{1/2}$. The iteration of $f(y)$ gives rise to the maximal magnification and overlapping factor $m=2$, $\mu = \ln 2$, and $R=1$, exactly.

Knowing the Lyapunov exponent μ_G for $G(y)$, we can compute the Lyapunov exponent $\mu(\lambda)$ for the original $f(x)$ at the $2N-N$ transition point for large N . In particular, for the $2^n \rightarrow 2^{n-1}$ transition at large n , we have

$$\mu(\lambda_n) = 2^{-n} \mu_G. \quad (4.28)$$

Using the asymptotic formula Eq. (3.6), we obtain

$$2^{-n} = \left(\frac{\lambda_n - \lambda_c}{A} \right)^t \quad (4.29)$$

with

$$t = \frac{\ln 2}{\ln 6} = 0.449\,806\,967. \quad (4.30)$$

Hence, we have

$$\mu(\lambda_n) = 0.6867 \left(\frac{\lambda_n - \lambda_c}{A} \right)^t. \quad (4.31)$$

The $(\lambda - \lambda_c)$ power dependence has been recently obtained by Huberman and Rudnick.³ We are also able to determine the prefactor in addition to the power dependence at these transition points.

V. BEHAVIOR OF EIGENVALUES NEAR THE CHAOTIC TRANSITIONS

We learned from Sec. II that there are two eigenvalues $\eta = \pm 1$ with $|\eta| = 1$ in the chaotic two-region. On the other hand, there is only one eigenvalue $\eta = 1$ with $|\eta| = 1$ in the chaotic one region. Thus, when the system makes a 2 to 1 chaotic transition, the $\eta = -1$ eigenvalue becomes an eigenvalue of $|\eta| < 1$.

In the chaotic two region, the eigenfunction ϕ_0 of $\eta = 1$ is real and positive. The support of ϕ_0 contains two separate regions I and II as in Fig. 4. The eigenfunction ϕ_1 associated with $\eta = -1$ also has the same support. We may choose ϕ_1 as

$$\phi_1 = \begin{cases} \phi_0, & \text{in region I} \\ -\phi_0, & \text{in region II.} \end{cases} \quad (5.1)$$

However, as one enters the chaotic one region, the previous two regions I and II overlap (see Fig. 5). To the first approximation, ϕ_0 in the overlapping region may be viewed as the sum of ϕ_0 's in I and II, while ϕ_1 in the overlapping region is the difference. The interference of ϕ 's in the overlapping region is responsible for reducing $|\eta|$ to less than 1. To understand qualitatively how η behaves near the transition, we consider the expression

$$\int dx dy K(x, y) \phi_0(y) \left(\frac{\phi(x)}{\phi_0(x)} + \frac{\phi(y)}{\phi_0(y)} \right)^2 = (2 + 2\eta) \int dx \frac{|\phi(x)|^2}{\phi_0(x)}, \quad (5.2)$$

which implies

$$1 + \eta = \frac{\int dx dy K(x, y) \phi_0(y) \left(\frac{\phi(x)}{\phi_0(x)} + \frac{\phi(y)}{\phi_0(y)} \right)^2}{2 \int dx \frac{|\phi(x)|^2}{\phi_0(x)}}. \quad (5.3)$$

In (5.2) and (5.3), we have made use of the fact that η and ϕ are real. When λ is smaller than the transition eigenvalue λ_1 , the regions I and II are disjoint. In this case, we have

$$K(x, y) \left(\frac{\phi(x)}{\phi_0(x)} + \frac{\phi(y)}{\phi_0(y)} \right)^2 = 0 \quad (5.4)$$

and hence

$$1 + \eta = 0, \quad \lambda < \lambda_1. \quad (5.5)$$

When $\lambda > \lambda_1$ regions I and II overlap and the left-hand side (lhs) of Eq. (5.2) no longer vanishes. We find that

$$\left(\frac{\phi(x)}{\phi_0(x)} + \frac{\phi(y)}{\phi_0(y)} \right)^2 = 0(1) \quad (5.6)$$

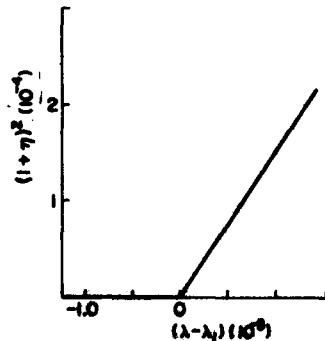


FIG. 15. $(1 + \eta)^2$ as a function of $\lambda - \lambda_1$ near the 2 to 1 chaotic transition $\lambda = 3.5706$.

for x in the overlapping region. One can show that the only important region of integration in the lhs of Eq. (5.2) is for x in the overlapping region. With x in the overlapping region (OR), we may have a crude estimate of $1 + \eta$ as,

$$1 + \eta \sim \frac{0(1) \int_{OR} dx \int dy K(x, y) \phi_0(y)}{\int dx \phi_0(x)} = \frac{0(1) \int_{OR} dx \phi_0(x)}{\int dx \phi_0(x)}. \quad (5.7)$$

In deriving (5.7), we have also approximated $|\phi(x)|^2 / \phi_0(x)$ in the denominator of Eq. (5.3) by ϕ_0 . Equation (5.7) indicates that $1 + \eta$ is proportional to the integrated density of ϕ_0 in the overlapping region. Note that the size of the overlapping region δ is proportional to $\delta\lambda = \lambda - \lambda_1$ and that $\phi_0(x)$ is dominated by square-root singularities $1/|x - x_c|^{1/2}$, in the overlapping region. Hence, we have

$$1 + \eta = \begin{cases} \propto \int \frac{dx}{|x - x_c|^{1/2}} \propto \sqrt{\delta\lambda}, & \lambda > \lambda_1 \\ 0, & \lambda < \lambda_1. \end{cases} \quad (5.8)$$

The constant of proportionality depends on the detailed structure of $\phi_0(x)$ and on the rate of change of the overlapping region. However, we know that in the vicinity of $\lambda = \lambda_1$, $(1 + \eta)^2$ is zero for $\lambda < \lambda_1$ and increases linearly with λ for $\lambda > \lambda_1$. In Fig. 15, we plot $(1 + \eta)^2$ as a function of λ . The transition and the linear dependence in λ are indeed verified numerically.

We may generalize the method to study the change of eigenvalue η near a 3-1 chaotic transition. In the chaotic three region, the eigenvalues with $|\eta| = 1$ are $1, \eta_0 = \frac{1}{2}(-1 + \sqrt{3}i)$, and η_0^* . The $\phi_0(\eta = 1)$ state contains three separate pieces I, II, and III as shown in Fig. 6. The η_0 (or η_0^*) state is obtained from ϕ_0 by assigning three different phases $1, \eta_0, \eta_0^*$ ($1, \eta_0^*, \eta_0$) to regions I, II, and III. In the chaotic one region, only the eigenvalue $\eta = 1$ remains to be of magnitude 1. The other two eigenvalues η, η^* now have magnitude less than one (i.e., $|\eta| < 1$). We can also study the rate of $|\eta|$ approaching one near the transition region $\lambda = \lambda_1$. In analogy to Eq. (5.3), we have

$$2 - \eta_0^* \eta - \eta_0 \eta^* = \frac{\int dx dy K(x, y) \phi_0(y) \left| \frac{\phi(x)}{\phi_0(x)} - \eta_0^* \frac{\phi(y)}{\phi_0(y)} \right|^2}{\int dx \frac{|\phi(x)|^2}{\phi_0(x)}}. \quad (5.9)$$

For $\lambda < \lambda_1$, we are in the chaotic three region and both sides of Eq. (5.9) vanish identically. For $\lambda > \lambda_1$ we are in the single chaotic region. As we have shown in Sec. IV, the gaps between regions

I and II and regions II and III develop nonvanishing ϕ_0 . The magnitude of ϕ_0 in the gaps is proportional to the incoming flux (source) via the gate DE . The magnitude of the source is given by

$$\text{Source} = \int_{\text{gap}} dx \phi_0(x) \propto \sqrt{\delta\lambda} \quad (5.10)$$

with $\delta\lambda = \lambda - \lambda_1$. Thus, $\phi_0(x)$ in the gap region is also proportional to $\sqrt{\delta\lambda}$. Now, we look at the numerator on the right-hand side (rhs) of Eq. (5.9). The important contribution to the numerator comes from the integration region where y is in the gap region DF . The requirement of $K(x, y) \neq 0$ implies that x is in the gaps or in region II. Then, the phase correlation factor is

$$\left| \frac{\phi(x)}{\phi_0(x)} - \eta_0^* \frac{\phi(y)}{\phi_0(y)} \right|^2 = 0(1).$$

We then have the estimate

$$\begin{aligned} 2 - \eta_0^* \eta - \eta_0 \eta^* &\approx 0(1) \frac{\int dx \int_{\text{gap}} dy K(x, y) \phi_0(y)}{\int dx \phi_0(x)} \\ &= 0(1) \int_{\text{gap}} dy \phi_0(y) / \int dx \phi_0(x) \propto \sqrt{\delta\lambda}. \end{aligned} \quad (5.11)$$

Near $\eta = \eta_0$, we can rewrite the lhs of (5.11) as

$$\begin{aligned} 2 - \eta_0^* \eta - \eta_0 \eta^* &= 1 - \eta \eta^* + |\eta_0 - \eta|^2 \\ &= 2(1 - |\eta|) + |\eta_0 - \eta|^2 - (1 - |\eta|)^2. \end{aligned} \quad (5.12)$$

Ignoring terms second order in smallness, we have

$$1 - |\eta| = \begin{cases} \propto \sqrt{\delta\lambda}, & \lambda > \lambda_1 \\ 0, & \lambda < \lambda_1 \end{cases} \quad (5.13)$$

as before. We have studied $(1 - |\eta|)^2$ as a function of λ numerically and indeed observed the above dependence.

ACKNOWLEDGMENTS

We wish to thank Professor A. Jackson and Professor R. L. Schult for stimulating discussions. This work was supported in part by the National Science Foundation under Grant No. NSF PHY 79-00272, and by the Office of Naval Research under Contract No. N00014-80-C-0840.

APPENDIX A: NUMERICAL METHOD

To find the eigenvalues and eigenfunctions of the operator K given in Eq. (3.7) we used a relatively straightforward, but possibly inefficient method. Basis functions for the space were chosen to be

$$f_n(x) = \begin{cases} 1, & x_n < x < x_{n+1} \\ 0, & \text{otherwise} \end{cases} \quad (A1)$$

where the x_n were a prescribed set of points on $(0, 1)$. K is then expanded in this basis set. The problem that arises is that a large matrix is needed (on the order of 1000×1000) for reasonable accuracy. The reason for this is essentially that discussed in Sec. II; that is, the width of K controls the number of bifurcations that are possible. A better choice of basis functions might alleviate this problem somewhat. Fortunately, most of the elements of K are zero, in fact, the number of nonzero elements is about twice the linear dimension of K .

We are now faced with finding the eigenvalues and eigenvectors of a large sparse matrix. Since we are only interested in the largest eigenvalues it is appropriate to use an iterative technique. To find the equilibrium-density distribution straightforward iteration of the discretized version of Eq. (A1) works fairly well, the only complication being in sorting out ϕ_0 from the other eigenvectors with eigenvalues of modulus 1. Our pictures of ϕ_0 used this method.

The eigenvalue (and associated eigenvector) problem is more complicated. We used a version of the Lanczos method. See, for example, Wilkinson⁷ for a complete description. The idea is to start from an arbitrary left state L_1 and an arbitrary right state R_1 and iterate:

$$\begin{aligned} KR_1 &= a_1 R_1 + b_2 R_2, \\ L_1 K &= a_1 L_1 + s_2 b_2 L_2. \end{aligned} \quad (A2)$$

At the n th step we have,

$$\begin{aligned} KR_n &= a_n R_n + b_n R_{n+1} + s_n b_n R_{n-1}, \\ L_n K &= a_n L_n + s_{n+1} b_{n+1} L_{n+1} + b_n L_{n-1}. \end{aligned} \quad (A3)$$

The newly generated R_{n+1} is chosen to be orthogonal to all previous generated L_n ; similarly for L . We choose $\langle L_n | R_n \rangle = 1$. This completely determines the coefficients a_n and b_n . s_n is just a sign ± 1 . In this scheme it is easily seen that a tridiagonal matrix for K is generated. We typically used a 30×30 matrix. Actually it is important to use several sizes say 26-30 and only the stable eigenvalues are relevant. That is, new eigenvalues are introduced at each stage and they typically change as the matrix size is increased. The eigenvectors can easily be calculated this way, but we have not actually done that. The absolute accuracy of predicting transitions was of the order of several percent, but the relative accuracy is much better. This enables us to compute the scaling of eigenvalues near transitions quite accurately.

APPENDIX B: ALTERNATIVE PROOF OF LEMMA 3

In this appendix we present an alternative proof of Lemma 3 that is valid for any number of dimensions. We start with a proposition that is proved in Ref. 6. It states that under fairly general conditions an operator satisfying Eqs. (2.2) and (2.3) has a cyclic peripheral spectrum. That is to say that if there are N eigenvalues with absolute value one, those eigenvalues are all of the N th roots of unity. If there is a degeneracy the spectrum must be divided into classes satisfying the above condition. We will assume that there is no degeneracy. Thus we are given

$$\phi_0, \phi_1, \dots, \phi_{N-1},$$

such that

$$K\phi_j = \omega^j \phi_j,$$

with

$$\omega = e^{2\pi i/N}. \quad (B1)$$

From Lemma 2 we know that we can choose

$$\phi_j(x) = e^{i\theta_j(x)} \phi_0(x), \quad j=1, \dots, N-1. \quad (B2)$$

We further choose the phases such that at a particular point $x=a$ with $\phi_0(a) \neq 0$,

$$\theta_j(a) = 0, \quad j=1, \dots, N-1. \quad (B3)$$

We first prove a Lemma: If, for K , satisfying Eqs. (2.2) and (2.3) there exist $\phi_0, \dots, \phi_{N-1}$ such that

$$\int K(x, y) \phi_j(y) dy = \phi_{j+1}(x) \quad (B4)$$

and

$$\phi_N(x) = \phi_0(x),$$

then

$$\int K(x, y) |\phi_j(y)| dy = |\phi_{j+1}(x)|. \quad (B5)$$

Proof. Taking the absolute value of Eq. (B4), we obtain

$$\int K(x, y) |\phi_j(y)| dy \geq |\phi_{j+1}(x)|. \quad (B6)$$

Integrating Eq. (B6) gives the cyclic set of inequalities

$$\int |\phi_j(x)| dx \geq \int |\phi_{j+1}(x)| dx.$$

Because of the cyclic nature the equality sign must hold, which implies Eq. (B5), except perhaps on a set of measure zero. Thus the Lemma is

proved.

We now prove the important theorem which says that the support of the eigenfunctions consist of disjoint regions and the phases of each eigenfunction are constant within those regions.

Theorem. There exist N nonnegative functions $\chi_j(x)$ having disjoint support such that

$$K\chi_j = \chi_{j+1} \quad (B7)$$

and the χ_j serve as a basis for expansion of the eigenfunctions having modulus 1.

Proof. Construct

$$\chi_0 = \frac{1}{N} |1 + e^{i\theta_1} + \dots + e^{i\theta_{N-1}}| \phi_0,$$

$$\chi_1 = \frac{1}{N} |1 + \omega e^{i\theta_1} + \omega^2 e^{i\theta_2} + \dots| \phi_0,$$

$$\chi_2 = \frac{1}{N} |1 + \omega^2 e^{i\theta_1} + \omega^4 e^{i\theta_2} + \dots| \phi_0,$$

...

$$\chi_{N-1} = \frac{1}{N} |1 + \omega^{N-1} e^{i\theta_1} + \dots| \phi_0.$$

It is trivial to verify from Eqs. (B1) and (B2) that without the absolute value signs, the χ_j satisfy Eq. (B4). From the above Lemma we see that they satisfy Eq. (B7). We further note that at $x=a$ where $\theta_j(a) = 0$,

$$\chi_0(a) = \phi_0 \neq 0,$$

$$\chi_j(a) = 0, \quad j \neq 0.$$

(This follows from the fact that the sum of the N th roots of unity is zero.) Next construct

$$\xi_0 = (\chi_0 + \chi_1 + \dots + \chi_{N-1}),$$

$$\xi_1 = \left(\chi_0 + \frac{1}{\omega} \chi_1 + \frac{1}{\omega^2} \chi_2 + \dots \right), \quad (B8)$$

$$\xi_2 = \left(\chi_0 + \frac{1}{\omega^2} \chi_1 + \dots \right),$$

etc. These functions satisfy

$$K\xi_j = \omega^j \xi_j,$$

so we conclude that

$$\xi_j = C_j \phi_j.$$

Now at $x=a$ we have

$$\xi_j = \phi_0, \quad j=0, 1, \dots, N-1.$$

Thus

$$C_j = 1, \quad j=0, \dots, N-1.$$

Taking the absolute value of Eqs. (B8) we obtain

$$\chi_0 + \chi_1 + \dots + \chi_{N-1} = \phi_0 = |\xi_j|, \quad j=1, \dots, N-1.$$

Because the coefficients ω^j point in different dir-

ections in the complex plane, the functions χ_j must have disjoint support. Finally the construction of ξ_j demonstrate the constancy of their phases (ω^{-jk}) in the k th region.

¹R. H. May, *Nature* **261**, 459 (1976).

²M. J. Feigenbaum, *J. Stat. Phys.* **19**, 25 (1978); **21**, 669 (1979), *Phys. Lett.* **74A**, 375 (1979).

³B. A. Huberman and J. Rudnick, *Phys. Rev. Lett.* **45**, 154 (1980).

⁴J. Crutchfield and B. Huberman (unpublished).

⁵J. Crutchfield, D. Farmer, N. Packard, R. Shaw,

G. Jones, and R. J. Donnelly, *Phys. Lett.* **76A**, 1 (1980).

⁶H. H. Schaefer, *Banach Lattices and Positive Operators* (Springer, Berlin, 1974), see Sec. (V.4).

⁷J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).

Iterative properties of a one-dimensional quartic map: Critical lines and tricritical behavior

Shau-Jin Chang, Michael Wortis, and Jon A. Wright

Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801

(Received 13 July 1981)

This paper studies the iterative properties of the two-parameter family of maps $x_{n+1} = 1 + ax_n^2 + bx_n^4$. These maps can have either one or three extrema. Multiple extrema lead to a complex region of iterative stability and to new types of iterative behavior. In particular, the Feigenbaum critical line (locus of 2^n -cycle accumulations) present for $b \approx 0$ terminates at two tricritical points T and T' with coordinates $(a, b) = (0, -1.59490)$ and $(-2.81403, 1.40701)$, respectively. There is also a "dual" critical line terminating in two additional tricritical points $\bar{T} = (-3.18980, 2.54371)$ and $\bar{T}' = (0.95561, -1.14981)$. Behavior near the tricritical points is controlled by a new fixed point f_7^* of the functional recursion relation $f_{n+1}(x) = [1/f_n(1)]^{1/2} f_n(x)$. This fixed point has two relevant directions and, therefore, behavior near a tricritical point depends on two new universal numbers $\delta_7^{(1)} = 7.28469$ and $\delta_7^{(2)} = 2.85713$. Crossover and scaling behavior near the tricritical points are explored.

I. INTRODUCTION

Iterative properties $x_{n+1} = f(x_n)$ of one-dimensional continuous maps provide what is probably the simplest examples of nonlinear dissipative dynamical behavior and have attracted considerable attention in recent years.^{1,2} Changing a single control parameter in such maps may lead from stable fixed-point behavior via a sequence of bifurcations to chaotic or quasichotic behavior. Feigenbaum³ showed that the limiting form of such bifurcation sequences exhibited scaling properties characterized by universal numbers, in strict analogy to critical behavior in statistical mechanics, and indeed, that the renormalization-group ideas developed in the statistical-mechanics context⁴ could be fruitfully carried over. Such bifurcation sequences are also seen in higher-dimensional models^{1,5,6} and in experiments.⁷ They represent one possible kind of "transition to chaos." Feigenbaum's universal numbers have, in fact, cropped up in some higher-dimensional systems.⁷

The prototypical one-dimensional map is the so-called logistic map,

$$f(x) = \lambda x(1-x), \quad (1.1)$$

which for $0 < \lambda < 4$ maps the unit interval $[0, 1]$ onto itself. Equation (1.1) is equivalent under change of origin and rescaling to the symmetric quadratic map,⁸

$$f(x) = 1 + ax^2. \quad (1.2)$$

In particular, the main bifurcation sequence has its limit at $a_c = -1.40116$. There is a simple sequence of 2^n cycles for $a_c < a < \frac{1}{2}$ and a corresponding inverse-bifurcation sequence of more complicated 2^n bands⁹ for $-2 < a < a_c$. The vicinity of the point a_c is characterized by a scaling behavior.³ For example, the position a_n of the superstable 2^n cycle for $a > a_c$ and large n varies as

$$a_n - a_c = A/\delta_p^n, \quad (1.3)$$

where $\delta_p = 4.66920$ is one of Feigenbaum's universal "critical exponents" and the nonuniversal amplitude $A = 9.349$. (The actual exponents in statistical mechanics are really related to $\ln \delta_p$.) So far, we have discussed only the logistic map (1.2). It is known, however, that the behavior of the logistic map provides a precise model for a large class¹⁰ of single-peak functions: The ordering of n cycles¹¹ and the critical behavior^{3,12} are universal; only the nonuniversal constants a_c and A vary from map to map.

It is the purpose of this paper¹³ to study the behavior of a simple two-parameter class of maps which includes (1.2), the symmetric quartic maps,

$$f(x) = 1 + ax^2 + bx^4. \quad (1.4)$$

The distinctive feature of the maps (1.4) is that for $ab < 0$ they have three extrema, at

$$x_0 = 0, \quad x_{\pm} = \pm \sqrt{-a/2b}, \quad (1.5)$$

rather than one as (1.2). This multiple-peak or multiple-valley structure makes possible a plethora of new, interesting, and complicated iterative behavior. It is the purpose of this paper to survey in a partially empirical and certainly nonrigorous way some of the new behavior in the context of the two-parameter family (1.4).

Section II of the paper treats general, noncritical properties. In the one-extremum regions ($ab > 0$) behavior like that of the logistic map is observed, only the accumulation point a_c is drawn out into a "critical line." This critical line terminates in the regions $ab < 0$, where it is "pinched off" at tricritical points T and T' by the existence of behavior more complicated than a monotonic cascade of 2^n -cycle regions. When $ab < 0$, it is quite possible for two stable cycles (but apparently no more than two) to coexist. Superstable cycles

must pass through one of the extrema (1.5). For special values of (a, b) a single "doubly stable" cycle may use two of them. Tricritical points are associated with $n \rightarrow \infty$ limits of doubly stable n cycles.

We develop a special form of topological conjugacy, called duality. Dual points P and \bar{P} have the same iterative properties. The dual of the critical line TT' is another critical line $\bar{T}\bar{T}'$ in the a, b plane.

We determine the region of the a, b plane over which the function (1.4) is iteratively stable, i.e., possesses a set of points x of nonzero measure with bounded itineraries.

Section III is devoted to the special properties of the map (1.4) near the endpoints T and T' of the critical line. Feigenbaum³ showed that the points on the critical line were associated with a fixed point of a certain functional recursion relation. The tricritical points T and T' are associated with a different fixed point of the same functional recursion relation. At this new fixed point there are two relevant directions and, therefore, two new universal numbers.

We discuss the scaling properties of the regions near the tricritical points and the crossover behavior between critical and tricritical regimes. The fixed points associated with $\bar{T}\bar{T}'$ are different from those associated with TT' but related to them by topological conjugacy.

II. GENERAL PROPERTIES UNDER ITERATION OF THE SYMMETRIC QUARTIC MAP

This section describes certain general, noncritical properties of the map (1.4).

A. Preliminaries

The possible shapes of the symmetric quartic map (1.4) depend on the relative signs and magnitudes of the parameters a and b , as shown in Fig. 1. The curves have one local extremum in quadrants I and III and three in quadrants II and IV. The number of intersections of $y=f(x)$ with the straight line $y=x$ may be zero, two, or four. The regions of each behavior are indicated in Fig. 1 and the corresponding curves are sketched in Fig. 2.

Itineraries $x = x_1, x_2(x), x_3(x), \dots$ of an initial point x under iteration (1.4) are of several types.² We refer to as bounded those itineraries such that $|x_n| \neq \infty$ as $n \rightarrow \infty$. Itineraries such that $|x_n| \rightarrow \infty$ as $n \rightarrow \infty$ are unbounded. The set of x having bounded itineraries is the iterative domain of f . When the iterative domain of f has nonzero measure, then f is iteratively stable, otherwise, it is iteratively unstable. We wish to explore in the

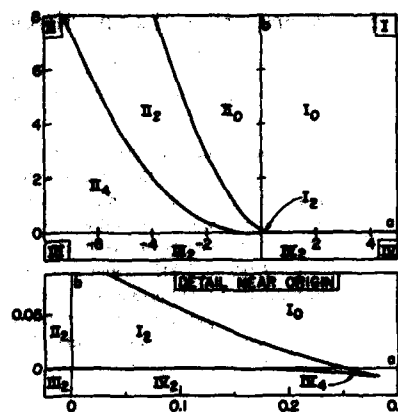


FIG. 1. Four quadrants of the a, b plane, indicating the number of intersections of $y=1+ax^2+bx^4$ with $y=x$. Detail shows region near the origin. Note that there is a small region of quadrant I where there are two intersections and an even smaller region of quadrant IV where there are four. Corresponding curves are sketched in Fig. 2.

remainder of this section the character and extent of the iterative domain of the map (1.4) for various values of the parameters a and b .

We distinguish the following types of bounded itineraries: (a) fixed points and other n cycles, (b) itineraries which, as $n \rightarrow \infty$, approach (are attracted by) a fixed point or other n cycle, and (c) other bounded itineraries, which we call chaotic.¹⁴ The set of n cycles corresponds to the set of solutions of

$$x = f^{(n)}(x), \quad (2.1)$$

where we have denoted $f^{(n)} = f \circ f \circ f \circ \dots \circ f$ and $f \circ g(x) = f(g(x))$. This set is countable and, therefore, contributes in and of itself zero measure to the iterative domain of f . The n cycle $x_1, x_2, \dots, x_n, x_{n+1} = x_1$ is stable, if a finite region of nearby x 's is attracted to it, otherwise, it is unstable.

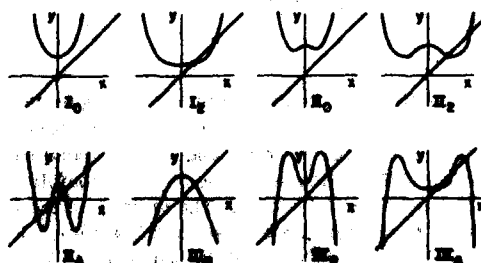


FIG. 2. Sketched shapes of the curve $y=1+ax^2+bx^4$ in the various regions of the a, b plane shown in Fig. 1. Intersections with $y=x$ are shown.

In particular, it is stable (unstable) 1.

$$\left| \frac{d}{dx_1} f^{(n)}(x_1) \right| < 1 \text{ (}> 1). \quad (2.2)$$

When

$$\frac{d}{dx_1} f^{(n)}(x_1) = 0, \quad (2.3)$$

the n cycle is *superstable*.¹⁵ A function f is iteratively stable if and only if it has one or more stable n cycles and/or a domain of x 's of nonzero measure with chaotic itineraries.

When f is iteratively stable, we shall wish to distinguish the number of different types of (bounded) limiting behavior occurring in the iterative domain of f . If all x 's except for a set of measure zero have the same limiting behavior (e.g., if all are attracted to a single stable n cycle or if all have chaotic itineraries using the same domain), then f is *unimodal*¹⁶; if two distinct types of limiting behavior occur with nonzero measure, then f is *bimodal*, and so forth.

B. Topological conjugacy duality

It is useful before proceeding further to establish two important symmetry properties. Let h be an arbitrary locally invertible map and define

$$g = h^{-1} \circ f \circ h, \quad (2.4)$$

so $f = h \circ g \circ h^{-1}$. Maps f and g related by (2.4) are called *topologically conjugate*.¹⁷ Itineraries of f and g are related to one another in a simple one-to-one manner: any bounded (unbounded) itinerary of f maps into a bounded (unbounded) itinerary of g and conversely. Map g is iteratively stable if and only if f is. Similarly, n cycles map to n cycles, chaotic itineraries to chaotic itineraries, etc.

For general h the functional forms of the topologically conjugate maps f and g will be quite different. Two instances when this is not so are translations,

$$h_1(a, x) = x - a, \quad h_1^{-1}(a, x) = x + a, \quad (2.5)$$

and scale changes,

$$h_2(a, x) = ax, \quad h_2^{-1}(a, x) = x/a, \quad (2.6)$$

both of which take n th degree polynomials into n th degree polynomials. For example, the general symmetric quartic polynomial ($C \neq 0$)

$$F(x) = C + Ax^2 + Bx^4 \quad (2.7)$$

is topologically conjugate under scale change to the standard, normalized form (1.4), with $a = AC$, $b = BC^3$ via

$$f = A_0^{-2}(C) \circ F \circ h_2(C), \quad (2.8)$$

since

$$F(x) = C \left[1 + AC \left(\frac{x}{C} \right)^2 + BC^3 \left(\frac{x}{C} \right)^4 \right]. \quad (2.9)$$

There is a special form of topological conjugacy which is exceedingly valuable in studying the iterative properties of the quartic map (1.4). Note that

$$f(x) = 1 - \frac{a^2}{4b} + b \left(\frac{a}{2b} + x^2 \right)^2 = G_1(G_2(x)), \quad (2.10)$$

with

$$G_1(x) = 1 - \frac{a^2}{4b} + bx^2$$

and

$$G_2(x) = \frac{a}{2b} + x^2. \quad (2.11)$$

Because $G_2 \circ G_1 = G_1^{-1} \circ (G_1 \circ G_2) \circ G_1$, f is topologically conjugate to the map $G_2 \circ G_1$ and, therefore, through (2.8) to the normalized map

$$\tilde{f} = h_s^{-1}(d) \circ G_2 \circ G_1 \circ h_s(d), \quad d = \frac{a}{2b} + \left(1 - \frac{a^2}{4b} \right)^2. \quad (2.12)$$

The map \tilde{f} is called the *dual* of f and has the form

$$\tilde{f}(x) = 1 + \tilde{a}x^2 + \tilde{b}x^4,$$

with

$$\begin{aligned} \tilde{a} &= (4b - a^2)[8ab + (4b - a^2)^2] / 32b^2, \\ \tilde{b} &= [8ab + (4b - a^2)^2] / (8b)^4. \end{aligned} \quad (2.13)$$

Duality is reciprocal, $(\tilde{\tilde{f}}) = f$. All of the topological correspondences mentioned after (2.4) hold between f and its dual. For example, if f has an n cycle using the point x , then \tilde{f} has a corresponding n cycle using the point

$$\tilde{x} = h^{-1}(d) \circ G_2(x) = \left(\frac{a}{2b} + x^2 \right) / d. \quad (2.14)$$

In particular, it follows from (2.14) that, if f has an n cycle using a side extremum ($x_1 = \pm \sqrt{-a/2b}$, $ab < 0$), then the corresponding n cycle of \tilde{f} uses the central extremum ($\tilde{x}_1 = 0$) and, conversely,¹⁷ if the n cycle of f uses the central extremum $x_0 = 0$, then that of \tilde{f} uses one of the side extrema $[\tilde{x}_0 = \pm (-3/2b)^{1/2}]$.

The duality relations (2.13) will be useful in Secs. IIC and IID below. The points on the curve

$$8ab + (4b - a^2)^2 = 0 \quad (2.15)$$

are all the dual to the origin. Remaining a axis points ($b = 0$, $a \neq 0$) map to infinity under duality, while remaining b -axis points ($a = 0$, $b \neq 0$) map to the parabola $b = a^2/4$. In addition, there is a line of self-duality, $(\tilde{a}, \tilde{b}) = (a, b)$, which can be param-

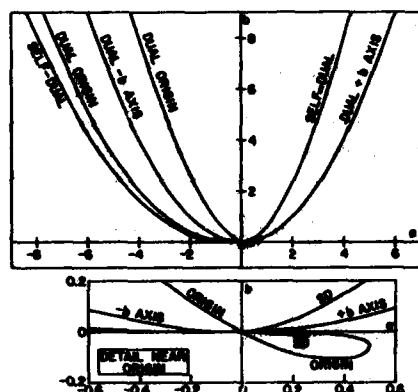


FIG. 3. Mapping of special lines under duality (2.13). The origin maps to the line (2.15). The b axis maps to the parabola $b = a^2/4$. Both the self-dual line and the inverse origin have branches in quadrant IV, as shown in the detail.

rized as

$$a = 2c^2(1+c), \quad b = c^2(1+c)^3, \quad (2.16)$$

corresponding to

$$f = h_s^{-1}(1+c) \circ g^{(2)} \circ h_s(1+c), \quad (2.17)$$

with

$$g(x) = 1 + cx^2.$$

These special lines are shown in Fig. 3. The mapping under duality of some sample regions of the a, b plane is shown in Fig. 4. Notice that regions A and C of Fig. 4 are adjacent, while their duals A and C meet only at the origin. A similar

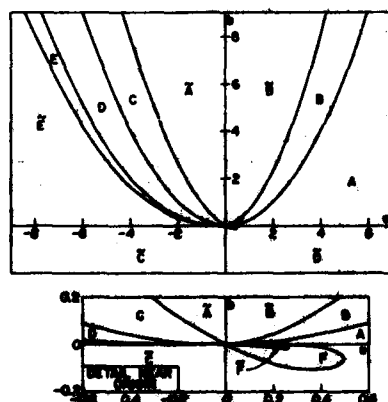


FIG. 4. Mapping of various regions under duality. Region A is the dual of region A , etc. Regions that are contiguous across the dual-origin line (2.15) do not remain so under duality.

remark applies to D and G and their duals \bar{D} and \bar{E} . This occurs because the origin is the dual image of the entire line (2.15).

The iterative behavior of f is completely known along the self-dual line, since $g(x)$ in (2.17) is the well-studied quadratic map (1.2). The map f is iteratively stable for all $-2 < c < \frac{1}{4}$ and otherwise unstable. At c values for which g has a stable $2n$ cycle, x_1, x_2, \dots, x_{2n} , f has two stable n cycles x_1, x_2, \dots and x_{2n}, x_{2n+1}, \dots , and is, therefore, bi-modal. At c values for which g has a stable $(2n+1)$ -cycle, f has a stable $(2n+1)$ cycle and is unimodal. The main bifurcation sequence of g ends at $c = -1.46116\dots$, which translates into $(a, b) = (-1.57518, 0.17758)$.

C. Stable, superstable, and doubly stable n cycles (Ref. 15)

It is straightforward to solve the n -cycle condition (2.1) numerically and then to check stability (2.2). The condition (2.3) for a superstable n cycle requires that one or more of the points x_1, x_2, \dots, x_n of the cycle be an extremum of f . The central extremum $x_0 = 0$ may always appear; however, at most one of the two side extrema can appear in any given cycle, since the evenness of f guarantees that the itineraries of $x_{\pm} = \pm \sqrt{-a/2b}$ ($ab < 0$) coincide after one step. Thus, each superstable cycle can be classified $+$, $-$, 0 , $0+$, or $0-$, according as it uses the indicated extrema. The last two cases denote n cycles ($n > 2$) which use both the central extremum and one side extremum and are, therefore, doubly stable. Doubly stable cycles are, indeed, particularly stable. When (1.4) has more than one extremum, all are quadratic. Thus, if x is a member of a doubly stable n cycle,

$$f^{(n)}(x + \delta x) = x + O(\delta x^2), \quad n = 4 \quad (2.18)$$

instead of $n = 2$ (ordinary superstable cycle). The argument following (2.14) shows that the dual of a doubly stable cycle is also doubly stable, provided only that the dual points \bar{x}_0 and \bar{x}_{\pm} are distinct. The exceptional case, $\bar{x}_0 = \bar{x}_{\pm} = 0$, occurs only for $\bar{a} = 0$, i.e., for

$$\bar{f}(x) = 1 + \bar{b}x^4, \quad (2.19)$$

in which case \bar{f} has a single quartic extremum. Any cycle which uses this quartic extremum satisfies (2.18) and will also be called doubly stable.

Suppose that for a particular set of parameters (a, b) f has a doubly stable cycle. It is clear that itineraries belonging to all points x sufficiently close to any of the extrema x_0, x_{\pm} are attracted to such a cycle. It turns out empirically (see Sec. II D) that f is always unimodal. If f is doubly stable for (a, b) , then f is unimodal for (a, b) in

some neighborhood of (a_0, b_0) . Doubly stable points will play an important role in the tricritical behavior which terminates the Feigenbaum critical line.

We illustrate some of these notations by simple examples. Regions of stable and superstable fixed points (1 cycles) are displayed in Figs. 5 and 6. Since $f(0)=1$ by our choice of normalization (1.4), superstable fixed points can only occur for (a, b) values such that

$$x_+(a, b) = x_+ = f(x_+) \text{ or } x_-(a, b) = x_- = f(x_-), \quad (2.20)$$

and the various branches of the curve in Fig. 5 have been labeled accordingly. The condition (2.20) reduces to (2.15). This is not an accident: Had we chosen the more general normalization (2.7), the condition for a superstable fixed point associated with the central peak would have been $C=0$, which via (2.8) is topologically conjugate to $a=b=0$. The discussion following (2.14) then shows that the x_+ superstable fixed points must occur for (a, b) dual to $(0, 0)$.

The regions of fixed-point stability shown in Figs. 5 and 6 contain the superstable line and are bounded [see (2.2)] by curves determined from

$$x_1(a, b) = x_1 = f(x_1),$$

with

$$f'(x_1) = \frac{df}{dx} \bigg|_{x=x_1} = \pm 1. \quad (2.21)$$

Over most of the 1-cycle region, f has only one stable fixed point^{1a}; however, there is a narrow region visible in Fig. 6 in which there are two independent stable fixed points. A typical example of this behavior is shown in Fig. 7: The domains of attraction of the two stable fixed points are disjoint and, between them, exhaust the iterative do-

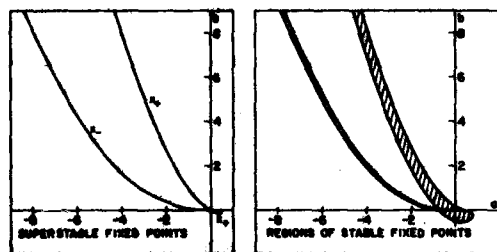


FIG. 5. Loci of stable and superstable fixed points in the a, b plane. The superstable locus coincides with the dual-origin curve of Fig. 3 and its branches have been labeled x_+ or x_- according to the associated extremum. The (cross-hatched) region of fixed-point stability includes the superstable locus. The upper edge of the stable region corresponds to the tangency conditions plotted in Fig. 1.

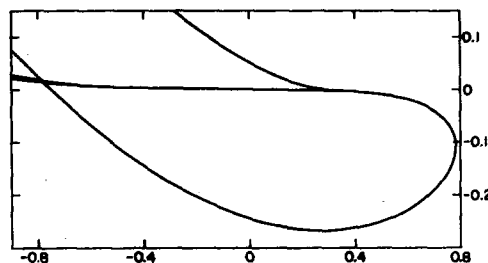


FIG. 6. Detail of the stable fixed-point region shown in Fig. 5. In the very narrow region of intersection of the two branches of stability the function $f(x)$ has two independent stable fixed points.

main of f (except, of course, for a set of measure zero, the unstable fixed points). The map f is, therefore, bimodal.

The superstable 2 cycles are shown in Fig. 8. The x_0 superstable 2 cycle satisfies

$$0 = x_0 = f^{(2)}(x_0) = f^{(2)}(0) = 1 + a + b. \quad (2.22)$$

The x_+ superstable 2 cycles can be found in analogy with (2.20); however, it is much easier to invoke duality and simply compute the dual of (2.22).

Intersections of superstable n cycle curves (e.g., Fig. 8) require comment. The set of curves belonging to a single extremum cannot intersect except at the origin. Similarly, curves belonging to x_+ cannot intersect those belonging to x_- , since x_+ and x_- have the same itinerary after one step. x_0 curves can intersect x_+ curves, as illustrated by Fig. 8. These intersections are of two quite dif-

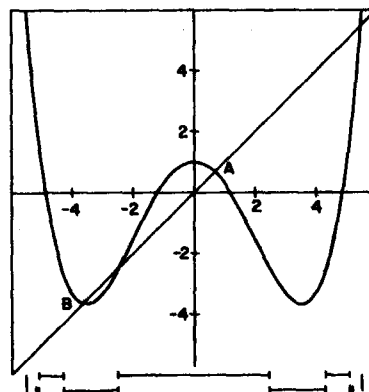


FIG. 7. Iterative plot of the function $f(x) = 1 - 0.75x^2 + 0.03x^4$. This function has two stable fixed points at $x_A = 0.670$ and $x_B = -3.662$. The iterative domain of f is $|x| < 5.475$. The domains of attraction of x_A and x_B are indicated below the graph. They exhaust the iterative domain of f (except for a set of measure zero) and interlace with each other as shown.

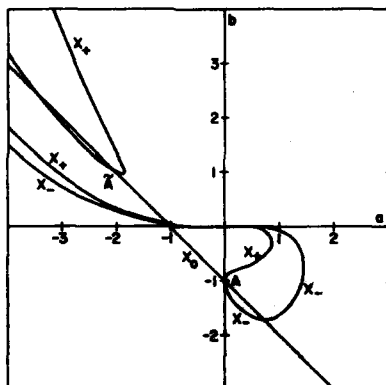


FIG. 8. Superstable 2 cycles. The central-extremum (x_0) curve is just the line $a+b+1=0$. Each branch is labeled according to the peak from which its stability derives. The point A and its dual \bar{A} are doubly stable. Other intersections are nondynamical.

ferent types. At the intersection point there may be two independent superstable cycles, each with its own domain of attraction. At such points f is bimodal and the curves are *dynamically independent*. Alternatively, there may be a single doubly stable cycle at the intersection point. In this case f is unimodal and the curves are *dynamically coupled*. In Fig. 8 only the intersection A and its dual image \bar{A} are dynamical. The remaining intersections are nondynamical.

Figure 9 shows the regions of stable 2-cycle behavior. Note the bimodal regions surrounding the nondynamical intersections of Fig. 8. The region of 1-cycle stability shown in Fig. 5 nests snugly against the region of 2-cycle stability, as shown in Fig. 10. In particular, following the

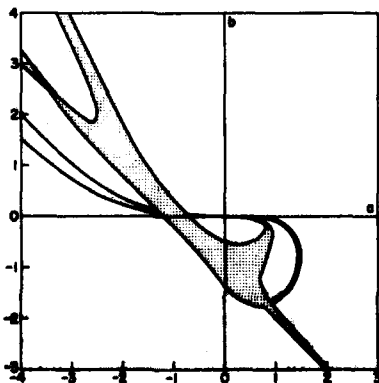


FIG. 9. Regions of stable 2-cycle behavior in the a, b plane.

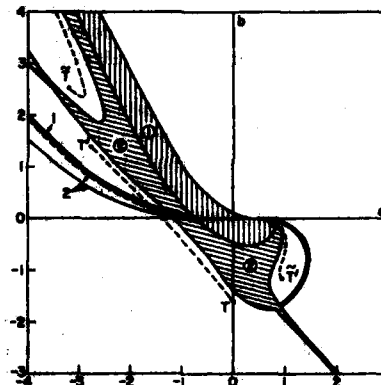


FIG. 10. Nesting regions of 1- and 2-cycle stability. Where such bifurcations proceed to completion, they end in a Feigenbaum critical line. The main critical line TT' and its dual $\bar{T}\bar{T}'$ are shown. The normal bifurcation sequence does not occur beyond these tricritical points.

negative a axis away from the origin, one sees the beginning of the main bifurcation sequence of the quadratic map (1.2). Such sequencing is also seen in other regions of Fig. 10. Whenever such bifurcations run to completion, they terminate in a Feigenbaum critical point. Regions of nesting therefore indicate a *Feigenbaum critical line*. Note, however, that the complicated topology of the 2-cycle regions already requires that the critical line be broken into segments. We have calculated the position of the main¹⁶ critical lines. They are included for convenience in Fig. 10 and will be discussed in Sec. III.

Figures 11 and 12 show loci of superstable 4 cycles and 3 cycles. Dynamical intersections (which will play a role in Sec. III) are indicated. The general picture which emerges for the central-extremum (x_0) curves is one or more diagonal lines running across the (a, b) plane from upper left to lower right and intersecting the negative a axis at the known values for the simple quadratic map (1.2) and in the standard sequence.²⁰ (Note that there are two such 4 cycles, one associated with bifurcation of the 2 cycle and the other in the region of inverse bifurcations on the far side of the critical line.) In addition, there are a large number of "fingers" in quadrants II and IV, some very narrow. These poke in towards the origin, cross no axes, and go off to infinity, apparently along the lines $b=0$, $a+b=0$, $a+b+1=0$, and $a+b+2=0$. The x_0 curves are obtained from the central-extremum curves by duality (2.13). Whenever an x_0 curve crosses (2.13) (see Fig. 3), the corresponding dual passes through the origin, creating the loops visible in quadrant IV.

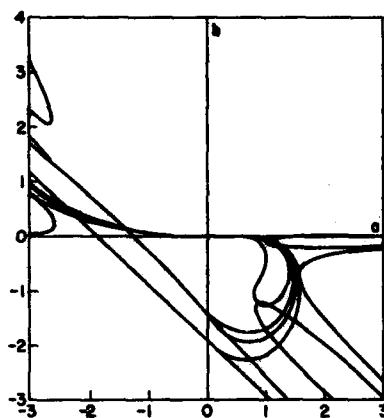


FIG. 11. Superstable 4 cycles. The 4 cycles necessary for the bifurcation sequences leading to TT' and $\bar{T}\bar{T}'$ (Fig. 10) is present. Much additional structure is also present. Several of the 4-cycle lines are too narrowly spaced to be resolvable on this graph.

D. Iterative stability

We now return to the question posed in Sec. II A of determining the region of the (a, b) plane over which the function (1.4) is iteratively stable. In principle, this is just the union over n of the regions of stability (see, e.g., Figs. 5 and 9, all n cycles plus, in addition, those regions where f has no stable n cycles but is iteratively stable by virtue of chaotic itineraries.²¹

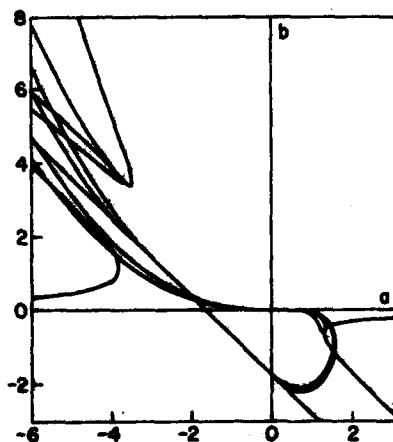


FIG. 12. Superstable 3 cycles. 3 cycles occur in the region beyond the main bifurcation sequence leading to TT' and $\bar{T}\bar{T}'$. Otherwise, 3-cycle structure is broadly similar to 2- and 4-cycle structures. Several of the 3-cycle lines are too narrowly spaced to be resolvable on this graph.

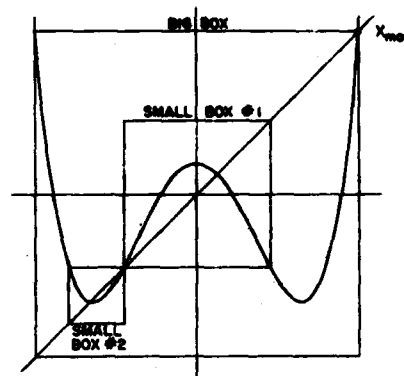


FIG. 13. Illustration of the box construction. A big box and two possible small boxes are shown. If the function remains inside any one (or more) of the boxes over the appropriate domain, then it is guaranteed to be iteratively stable.

Maps of the types I_0 and II_0 (see Figs. 1 and 2), for which there are no solutions of $x=f(x)$, have unbounded itineraries for all x . Their iterative domain is the empty set and they are iteratively unstable. At the boundary between $(I_0 + II_0)$ and $(I_2 + II_2)$, $y=f(x)$ is tangent to $y=x$, which is just the condition (2.21) giving the upper edge of the stable 1-cycle region. Just beyond this boundary in $(I_2 + II_2)$ there are two fixed points, x_{\max} (see Fig. 2) and x_{\min} , with $f'(x_{\max}) > 1$ and $f'(x_{\min}) < 1$. If $f'(x_{\min}) > -1$ (as is true throughout I_2), then any x with $|x| < x_{\max}$ flows to x_{\min} and any x with $|x| > x_{\max}$ has an unbounded itinerary, so f is iteratively stable and unimodal.

In other regions of the (a, b) plane the situation is more complicated and we must develop a methodology for establishing iterative stability. Figure 13 illustrates one such criterion, which we call the *box construction*. Let $x_{\max} = x_{\max}(a, b)$ denote the solution of $x=f(x)$ of largest absolute value. Suppose

$$|x_{\max}| \geq |f(x)| \quad \forall \quad |x| \leq |x_{\max}|. \quad (2.23)$$

We refer to this as the "big-box" condition. It is easy to see that points with $|x| \leq |x_{\max}|$ have bounded itineraries, while those with $|x| > |x_{\max}|$ have unbounded itineraries. Similar "small-box" constructions can be done separately for central and side extrema, as suggested in Fig. 13. If any or all of the extrema of f are boxable in the above sense, then f is iteratively stable. For example, in quadrant III, where f has a single peak at $x_0 = 0$, the box condition (2.23) requires

$$|x_{\max}| \geq f(0) = 1, \quad (2.24)$$

so the boundary of guaranteed stability is the line

$$-1 = x_{\max} = f(x_{\max}) = f(1) = 1 + a + b,$$

and (2.24) is equivalent to

$$a + b \geq -2. \quad (2.25)$$

When two peaks are independently (small-) boxable, then f is at least bimodal. Systematic application of the box conditions to all peaks and over the whole (a, b) plane leads to a region S of guaranteed iterative stability, pictured in Fig. 14. This rather complicated figure goes into itself under duality.

Failure of the box conditions may or may not imply iterative instability. For example, comparison of Figs. 9 and 14 already shows that the region of stable 2 cycles extends beyond the boundary of guaranteed stability into regions \tilde{D}_1 and \tilde{D}_2 of quadrants II and IV, respectively. We develop this situation in more detail. The map (1.4) is infinitely differentiable and it is easy to verify that its Schwartzian derivative

$$Sf(x) = \frac{f'''(x)}{f'(x)} - \frac{3}{2} \left(\frac{f''(x)}{f'(x)} \right)^2 \quad (2.26)$$

is everywhere negative. In quadrant III, where f is single peaked, it satisfies the hypotheses of a theorem due to Singer,^{2,23} which concludes that any stable n cycle attracts the peak. Thus, when (2.25) is satisfied (i.e., when the peak itinerary is bounded), f has at most a single stable n cycle, while, when (2.25) fails (i.e., when the peak itin-

erary is unbounded), f has no stable n cycles. In the former case f is iteratively stable and unimodal.²³ In the latter case f is iteratively unstable.²⁴ Thus, the region $a + b < -2$ of quadrant III and its dual image in quadrant II are iteratively unstable.

In quadrants II and IV, f has three extrema and does not satisfy the hypotheses of Singer's theorem. We have proceeded empirically. The result of our study may be summarized in a form that applies over the entire (a, b) plane by the following three statements.

(1) When all extrema have unbounded itineraries, then f is iteratively unstable.

(2) When $x_0 = 0$ has a bounded itinerary and x_1 have an unbounded itinerary or vice versa, then f is iteratively stable and unimodal.

(3) When both x_0 and x_1 have bounded itineraries, then f is iteratively stable. If the two itineraries are asymptotically distinct, then f is bimodal; if they are asymptotically convergent, then f is unimodal.

If all extrema lie outside the big box of Fig. 13, then statement (1) applies and f is iteratively unstable. Such regions have been indicated U (unstable) in Fig. 14. If all extrema lie inside the big box of Fig. 13, then statement (3) applies and f is iteratively stable. If the central extremum lies outside the big box and the x_1 extrema lie inside or vice versa, then there are two possibilities: If an inside extremum is small-boxable, then f is stable and unimodal [statement (2)], if not, then the situation remains in doubt. These doubtful regions are denoted D in Fig. 14. We investigated them numerically by the following method: We fix (a, b) and follow the itinerary of the "inside" extremum for 32 steps. If the itinerary ever reaches the region $|x| > |x_{\max}(a, b)|$, then it is unbounded and f is iteratively unstable. If the itinerary remains within the range $-|x_{\max}| < x < |x_{\max}|$, then we declare it bounded (for practical purposes) and f is iteratively stable and unimodal. In practice, the distinction is sharp. Figure 15 shows the results of a scan in (a, b) in region D_1 at a grid interval of 0.015. Similar structure occurs in the other D regions. What appears is a complex set of spines or *spicules* radiating outward from the known-stable region S . The major spicules are readily identifiable with low- n regions of n -cycle stability, which we have seen in Figs. 9, 12, and 13. Many other spicules certainly exist in these regions, however, they become rapidly more narrow as n increases and slip between the mesh of our grid scan. Well-defined regions of entirely unstable behavior are also suggested by Fig. 15. The boundary between

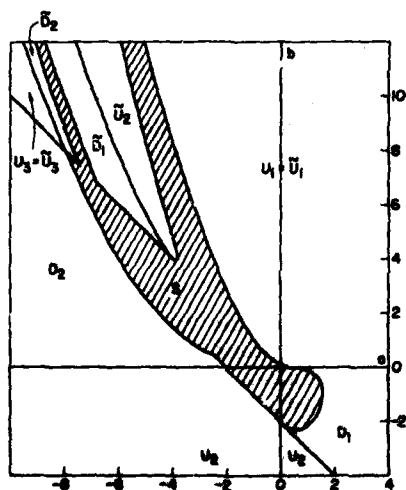


FIG. 14. Iterative stability via the box construction. The cross-hatched region S is guaranteed stable. The regions U and their duals \tilde{U} are guaranteed unstable. Stability of the regions D and their duals \tilde{D} cannot be decided by the simple box construction.

AD-A119 720

ILLINOIS UNIV AT URBANA DEPT OF PHYSICS
THEORETICAL STUDIES OF THE OCEANIC INTERNAL WAVE SYSTEM.(U)
AUG 82 J WRIGHT

F/G 8/3

N00014-80-C-0840

NL

UNCLASSIFIED

2 2

2

END
11-82

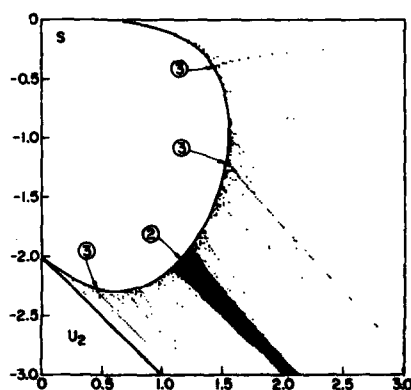


FIG. 15. Iterative stability in the doubtful region D_1 of Fig. 14. Regions S and U_2 are guaranteed stable and unstable, respectively, by the box construction. Points shown were determined iteratively stable by a 200×200 scan, as described in the text. The major splines are readily identifiable with low- n ($n=2,3$) n cycles, as indicated.

stable and unstable regions appears to be very irregular and may have some fractal character. All the above considerations could be further refined by studying box constructions for the functions $f^{(n)}$ for $n > 1$.

III. CRITICAL AND TRICRITICAL BEHAVIOR

A. The Feigenbaum critical line and its dual

The nesting 1- and 2-cycle regions shown in Fig. 10 are the first stages of a standard bifurcation cascade, 1, 2, 4, 8, When this cascade proceeds smoothly to completion, it terminates at a locus of accumulation points, which we shall call a critical line.⁴ We already know two points on the critical line: $(a,b) = (-1.40116, 0)$ on the negative- a (logistic) axis and the intersection of the critical line and the self-dual line given at the end of Sec. II B. There is no difficulty in finding the rest of the critical line numerically. The most efficient method is simply to plot the appropriate loci of central-peak superstable 4 cycles (Fig. 11), 8 cycles, etc., which rapidly converge to the Feigenbaum critical line TT' , $a_n = a_n(b)$ [or $b_n = b_n(a)$], shown in Fig. 10. Approaching this line from the upper right always leads through the standard cascade; continuing beyond to the lower left reveals the usual sequence of inverse-bifurcation bands. Asymptotically close to the critical line the position of the superstable 2^n cycles (or any other reference feature one cares to choose) varies [c.f. (1.3)] as

$$a_n(b) = a_c(b) + A(b)/\delta_f^n,$$

or

$$b_n(a) = b_c(a) + C(a)/\delta_f^n, \quad (3.1)$$

where the critical amplitudes $A(b)$ or $C(a)$ vary smoothly along the critical line but δ_f is universal.

The new feature here is the fact that the critical line terminates at the special points

$$T = (0, -1.59490), \quad T' = (-2.81403, 1.40701), \quad (3.2)$$

which we shall refer to as *tricritical points* in analogy to the theory of thermodynamic phase transitions.²⁵ The mechanism for the termination is the existence of new n -cycle structures which pinch off the orderly sequencing characteristic of the critical line. A variety of interrelated structures exist, which we shall study in more detail in Sec. III C. For the moment we mention three such structures: (i) A tricritical point is the limit of a sequence of doubly stable 2^n -cycle points. For example, the points T and T' are the limit of the dynamically coupled points listed in Table I, the first two of which are visible in Figs. 8 and 11. (ii) A tricritical point is the limit of a sequence of dynamically independent 2^n -cycle intersections (the first few are also visible in Figs. 8 and 11). As a consequence of (i) and (ii) any neighborhood of a tricritical point contains an infinite sequence of interlaced unimodal regions (associated with the doubly stable points) and bimodal regions (associated with the dynamically independent intersections). (iii) Finally, a tricritical point is also characterized as a point at which the critical line is pinched off by a sequence of narrow fingers (see the end of Sec. II C) of high-order out-of-order 2^n cycles. Such a finger is already visible in the 4 cycles (Fig. 8) in quadrant IV, where it is still well away from T [the finger tip only reaches $(0.80, -1.25)$; however, the corresponding 32-cycle finger reaches $(0.03, -1.62)$ and higher-order cycles close in rapidly].

Thanks to Feigenbaum³ we know that the critical

TABLE I. Sequences of doubly stable 2^n -cycle points (a,b) converging to the tricritical points T and T' , which terminate the Feigenbaum critical line. The corresponding points for \tilde{T} and \tilde{T}' may be obtained by using the duality relations (2.13).

Cycle	T	T'
2	(0, -1)	(-2, 1)
4	(0, -1.50555)	(-2.00516, 1.34251)
8	(0, -1.50825)	(-2.73575, 1.39758)
16	(0, -1.50916)	(-2.81159, 1.40575)
32	(0, -1.59490)	(-2.81200, 1.40704)
∞	(0, -1.59490)	(-2.81403, 1.40701)

line is associated with a fixed point of the functional recursion relation,

$$f_{n+1}(x) = \frac{1}{f_n(1)} f_n^{(2)}(x f_n(1)),$$

i.e.,

$$f_{n+1} = h_s^{-1}(f_n(1)) \circ f_n^{(2)} \circ h_s(f_n(1)). \quad (3.3)$$

The factor $f_n(1)$ in (3.3) is chosen to preserve normalization, $f_{n+1}(0) = 1$ provided $f_n(0) = 1$. At any fixed point of (3.3) (and, as we shall see, there are many),

$$f^* = h_s(\alpha) \circ f^{*(2)} \circ h_s^{-1}(\alpha), \quad (3.4)$$

with

$$\alpha = 1/f^*(1), \quad (3.5)$$

i.e., f^* is topologically conjugate to $f^{*(2)}$ under rescaling by magnification factor α . The mapping (3.4) takes a function $f_n = f^* + \delta f$ near f^* into another function $f_{n+1} = f^* + \delta f'$ near f^* , with

$$\delta f' = \mathcal{L}(\delta f) + O(\delta f^2), \quad (3.6)$$

where \mathcal{L} is linear. The eigenvalues Λ_m and eigenvectors e_m of \mathcal{L} satisfy

$$\mathcal{L}(e_m) = \Lambda_m e_m \quad (3.7)$$

and provide a convenient basis for describing the region of function space near f^* . For example, if

$$\delta f = \sum_m a_m e_m \quad (3.8)$$

for some set of coefficients a_m , then

$$\delta f' = \sum_m a_m \Lambda_m e_m. \quad (3.9)$$

The fixed point (3.4) of (3.3) associated with the Feigenbaum critical line has the form²

$$f_f^*(x) = 1 + c_2^{(f)} x^2 + c_4^{(f)} x^4 + \dots, \quad (3.10)$$

with

$$c_2^{(f)} = -1.5276, \quad c_4^{(f)} = 0.1048, \dots,$$

and

$$\alpha_f = 2.50281. \quad (3.11)$$

There is only one "relevant"²² eigenvalue $\Lambda_1 = \delta_f = 4.66920$. All other eigenvalues are "irrelevant," $|\Lambda_m| < 1$ for $m > 1$, so the corresponding perturbations (3.8) shrink under iteration. The fixed point f_f^* therefore has an inflowing critical hypersurface of codimension one. The Feigenbaum critical line is the intersection of this critical hypersurface with the a, b plane. Any function f_a on the Feigenbaum critical line (but excluding the endpoints T and T') flows to f_f^* under iteration (3.3). [Nearby

functions flow near to f_f^* , through the linear region (3.8), (3.9), and then (3.1) follows.] In particular, if $f_0(x) = f(x)$, then iteration under (3.3) gives

$$f_n = h_s^{-1}(f^{(2^n)}(0)) \circ f^{(2^n)} \circ h_s(f^{(2^n)}(0)), \quad (3.12)$$

with

$$f^{(2^n)}(0) = \prod_{m=1}^{n-1} f_m(1). \quad (3.13)$$

So, if f_a is on the Feigenbaum critical line, then

$$\lim_{n \rightarrow \infty} (f_n)_a = f_f^*. \quad (3.14)$$

Equation (3.14) is, in fact, a convenient way of finding the fixed-point function f_f^* . In addition, putting $x = 1$ in (3.14) and using (3.12) and the definition (3.5) of the magnification factor, we find

$$\alpha_f = \lim_{n \rightarrow \infty} [f_0^{(2^n)}(0)/f_0^{(2^{n+1})}(0)], \quad (3.15)$$

which is numerically convenient.

The dual of the Feigenbaum critical line TT' is another critical line $\bar{T}\bar{T}'$, shown in Fig. 10. Endpoints of this dual line are tricritical points located dual to (3.2),

$$\begin{aligned} \bar{T} &= (-3.18980, 2.54371), \\ \bar{T}' &= (0.95561, -1.14981). \end{aligned} \quad (3.16)$$

The dual critical line passes through (0,0) because TT' crosses (2.15) and through infinity because TT' crosses the a axis. The critical lines TT' and $\bar{T}\bar{T}'$ intersect at the self-dual line; however, they are dynamically independent: Feigenbaum critical behavior uses the central extremum, while its dual uses the x_1 extrema, in accordance with the discussion after (2.14). The continuity of the duality relations (2.13) guarantees that (3.1) holds in the vicinity of the dual critical line. Does the dual line lie on the critical hypersurface of the fixed point f_f^* ? The answer is no. Indeed, the iteration procedure (3.2)–(3.4) does not even lead to a function with finite coefficients like (3.10). The reason is simple enough. The 2^n cycles in the bifurcation behavior of $\bar{T}\bar{T}'$ use one of the side extrema x_1 instead of the central extremum. However, the iteration (3.3) continually magnifies the function around the origin, pushing the active part of the function near the x_1 extrema off to infinity. To achieve a finite fixed point, it would be necessary to translate the active peak to the origin before applying (3.3), (3.12), and (3.14). The infinite fixed point which controls $\bar{T}\bar{T}'$ is topologically conjugate to f_f^* in this sense.

B. Tricriticality: Exponent

The tricritical point T is located as the limit of the set of doubly stable 2^n cycles situated on the

negative- b axis. Locating T' would appear to require a two-dimensional search. Luckily, this is not so, as we now explain. For each $n > 1$ there are several doubly stable points in quadrant II (see Sec. III D). Each such cycle uses the extrema x_0 and x_n ; however, the different doubly stable points use them in different orders. The sequence of doubly stable 2^n cycles which converges to T' turns out to use the two extrema in the order x_0, x_n, \dots . Normalization requires $f(x_0) = f(0) = 1 = x_n(a, b)$, so $a = -2b$, as exemplified in Table I. Thus, the doubly stable points leading to T' all lie on the line

$$f(x) = 1 - 2cx^2 + cx^4 \quad (3.17)$$

and can be found by a one-dimensional search.

We turn now to tricritical behavior and start with an analysis of the point T . Functions on the b axis have the form

$$f(x) = 1 + bx^4 \quad (3.18)$$

and are iteratively stable and unimodal for $-2 < b < \frac{1}{16}$. The iterative behavior of such functions as b travels down the negative- b axis is entirely similar to that of the functions (1.3): There is a main bifurcation sequence (the beginnings of which we have already seen in Figs. 10 and 11 and Table I), which has its limit at $b_T = -1.5949013862388 \dots$. For $-2 < b < b_T$, there is a corresponding sequence of inverse-bifurcation bands. The vicinity of b_T is characterized by scaling behavior. For example, the positions b_n of the doubly stable 2^n cycles for $b > b_T$ and large n vary as

$$b_n - b_T = B/(\delta_f^{(1)})^n, \quad (3.19)$$

with $\delta_f^{(1)} = 7.28469$ and $B = 4.886$, in precise parallelism to (1.3). The number $\delta_f^{(1)}$ differs from δ_f because (3.18) has a quartic rather than quadratic maximum³ but is otherwise universal in the same sense. The amplitude B is nonuniversal.

The origin of the scaling behavior (3.19) is the existence of a new fixed point of (3.3), which we denote [cf. (3.10)]

$$f \#(x) = 1 + c_1^{(T)}x^4 + c_2^{(T)}x^8 + \dots, \quad (3.20)$$

where $f \#$ satisfies (3.4) with magnification factor

$$\alpha_T = 1/f \#(1) = (1 + c_1^{(T)} + c_2^{(T)} + \dots)^{-1}. \quad (3.21)$$

There are several ways of finding the tricritical fixed-point function.²⁷ One of the simplest ways follows the logic of (3.12)–(3.15): when $b = b_T$, $f(x) = f_T(x) = 1 + b_T x^4$. Then

$$f \# = \lim_{n \rightarrow \infty} (f_T^n) \quad (3.22)$$

and

$$\alpha_T = \lim_{n \rightarrow \infty} [f_T^{(2^n)}(0)/f_T^{(2^{n+1})}(0)]. \quad (3.23)$$

Equation (3.22) can be used to find the coefficients $c_i^{(T)}$ as follows: Let

$$f_T^{(n)}(x) = f_T^{(n)}(0) + C_1^{(n)}x^4 + \dots, \quad (3.24)$$

then it is easy to develop the recursion relation,

$$C_1^{(n+1)} = 4b_T [f_T^{(n)}(0)]^2 C_1^{(n)}. \quad (3.25)$$

However, $C_1^{(1)} = b_T$, so

$$C_1^{(n)} = \frac{1}{4} (4b_T)^n \prod_{k=1}^{n-1} [f_T^{(k)}(0)]^2,$$

and inclusion of the additional factors in (3.12) yields

$$c_1^{(T)} = \lim_{n \rightarrow \infty} \frac{1}{4} (4b_T)^n \prod_{k=1}^{n-1} [f_T^{(k)}(0)]^2. \quad (3.26)$$

Derivation of similar expressions for higher-order coefficients is straightforward. We find from (3.23)

$$\alpha_T = -1.69030 \dots \quad (3.27)$$

Results for $C_i^{(T)}$ are given in Table II.

Perturbation about the fixed point (3.20) determines scaling behavior, as outlined in Sec. III A. There are two types of perturbations, $\delta f = \delta f^{(e)} + \delta f^{(o)}$, with

$$\delta f^{(e)} = a_e x^4 + a_o x^6 + \dots \quad (3.28)$$

and

$$\delta f^{(o)} = a_e x^2 + a_o x^4 + \dots, \quad (3.29)$$

involving powers x^{2b} with even or odd b , respectively. Because $f \#$ is a polynomial in x^4 , the iteration of an even δf cannot generate any odd component. Hence, in terms of $\delta f = (\delta f^{(e)}(x))$, we have

$$\begin{pmatrix} \delta f^{(e)}(x) \\ \delta f^{(o)}(x) \end{pmatrix} = \mathcal{L}(\delta f) = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \begin{pmatrix} \delta f^{(e)}(x) \\ \delta f^{(o)}(x) \end{pmatrix}.$$

TABLE II. Coefficients of the tricritical fixed-point function $f \#(x) = 1 + c_1^{(T)}x^4 + c_2^{(T)}x^8 + \dots$. These coefficients were obtained via (3.26) and its higher-order analogs. Note that the sum rule (3.21) is well satisfied.

$c_1^{(T)}$	-1.53411
$c_2^{(T)}$	0.01269
$c_3^{(T)}$	0.21190
$c_4^{(T)}$	-0.00202
$c_5^{(T)}$	0.00000
$c_6^{(T)}$	0.00000
$c_7^{(T)}$	0.00000
$c_8^{(T)}$	0.00000

Although $\delta f^{(e)}$ couples to both $\delta f^{(e)}$ and $\delta f^{(o)}$, $\delta f^{(o)}$ couples only to $\delta f^{(e)}$. The reducibility of (δf) into a semiblock form implies that the eigenvalue problems (3.7) for the even and odd perturbations do not mix and may be considered independently. Both even and odd subspaces have, of course, an infinite number of eigenvalues and eigenvectors; however, in each case only one eigenvalue is relevant ($|\Lambda| > 1$) and the rest are irrelevant ($|\Lambda| < 1$). The largest even eigenvalue controls flows in the b (x^4) direction near tricriticality and is just the universal number $\delta_f^{(2)}$ introduced in (3.19) and determined numerically from

$$f_{n+1}(x) = f \frac{1}{2}(x) + \delta f'(x) \\ = \frac{1}{f_n(1)} [f \frac{1}{2}(x) (x f_n(1)) + \delta f(x f_n(1)) f \frac{1}{2}(x f_n(1)) + \delta f(f \frac{1}{2}(x f_n(1))) + O(\delta f^2)]. \quad (3.31)$$

The first and third terms in the brackets are always even; only the second term can be odd. As we mentioned earlier, we only need to work in the odd k space to obtain the desired eigenvalue. Thus, to linear order, we have

$$\delta f^{(o)}(x) = a'_2 x^2 + a'_4 x^4 + \dots \\ = \frac{1}{f \frac{1}{2}(1)} f \frac{1}{2}(x f \frac{1}{2}(1)) \delta f^{(e)}(x f \frac{1}{2}(1)). \quad (3.32)$$

Equation (3.32) has the structure

$$a'_k = \sum_l M_{kl} (c^{(T)}) a_l, \quad k, l = 2, 6, 10, \dots \quad (3.33)$$

The eigenvalues of M are the odd eigenvalues of (3.7). Because the coefficient of $\delta f^{(e)}$ in (3.32) is a polynomial in x^4 , a'_k cannot depend on a_l for $l > k$, so M is triangular and its eigenvalues are simply M_{kk} . To find M_{kk} , we can set $x = 0$ in the coefficient of $\delta f^{(e)}$ to get

$$M_{kk} = [f \frac{1}{2}(1)]^{k-2} f \frac{1}{2}'(1). \quad (3.34)$$

However, $f \frac{1}{2}$ satisfies (3.4),

$$f \frac{1}{2}(x) = \frac{1}{f \frac{1}{2}(1)} f \frac{1}{2}(x) (x f \frac{1}{2}(1)). \quad (3.35)$$

Differentiating (3.35) with respect to x and then letting $x \rightarrow 0$ leads to

$$1 = f \frac{1}{2}'(1) [f \frac{1}{2}(1)]^2, \quad (3.36)$$

so finally,

$$M_{kk} = [f \frac{1}{2}(1)]^{k-2} = \alpha_f^{k-2}, \quad (3.37)$$

which gives (3.30) and the irrelevant odd eigenvalues.

Because there are two relevant eigenvalues at the fixed point $f \frac{1}{2}$, the tricritical hypersurface, which flows to $f \frac{1}{2}$ under iteration (3.3), has no-

superstable 2^n -cycle spacings. Finding it and the other even eigenvalues by solving (3.7) with (3.28) is feasible but uninteresting. The largest odd eigenvalues we denote $\delta_f^{(2)}$. We shall show below that $\delta_f^{(2)}$ is simply related to the magnification factor

$$\delta_f^{(2)} = \alpha_f^{-2} = 2.65713, \quad (3.38)$$

and that the irrelevant odd eigenvalues are just the inverse odd powers of (3.30), $(\delta_f^{(2)})^{-(2n+1)}$. To see this, take $f_n = f \frac{1}{2} + \delta f$ and iterate (3.3) once to get

dimension equal to 2. This hypersurface evidently intersects the a, b plane at T and controls the scaling behavior there (see Sec. III C). It is by no means obvious that the tricritical point T' at the other end of the Feigenbaum line is also on the tricritical hypersurface belonging to $f \frac{1}{2}$. We have verified numerically that it is by observing that $\lim_{n \rightarrow \infty} (f_n)_n = f \frac{1}{2}$ [via (3.23), (3.26), etc.]. It follows that behavior near T' is described by the eigenvalues $\delta_f^{(1)}$, $\delta_f^{(2)}$, and α_f .

We turn now to the tricritical points terminating the dual critical line $\tilde{T}\tilde{T}'$. Duality guarantees that scaling behavior near \tilde{T} and \tilde{T}' is described by the universal numbers $\delta_{\tilde{f}}^{(1)}$ and $\delta_{\tilde{f}}^{(2)}$. It does not follow, however, that \tilde{T} and \tilde{T}' flow to the fixed point $f \frac{1}{2}$ and, indeed, they do not (cf. the situation with the dual critical line). Direct iteration in analogy to (3.22)–(3.26) shows that \tilde{T} and \tilde{T}' flow to another tricritical fixed point, given by

$$f \frac{1}{2}(x) = [f \frac{1}{2}(\sqrt{x})]^2, \quad (3.39)$$

with

$$\alpha_{\tilde{f}} = \alpha_f^2. \quad (3.40)$$

Note that (3.39) is just of the form (2.4), with $h(x) = \sqrt{x}$, $h^{-1}(x) = x^2$, and so $f \frac{1}{2}$ is topologically conjugate to $f \frac{1}{2}$. These amusing relations are not hard to derive: When applied to points on the b axis, the duality relations (2.10) and (2.11) read

$$G_1(b) = 1 + bx^2, \quad G_2(b) = x^2, \quad (3.41)$$

so

$$f \frac{1}{2}(x) = f \frac{1}{2}(x) = G_2(G_1(x)) = (1 + bx^2)^2 = [f \frac{1}{2}(\sqrt{x})]^2. \quad (3.42)$$

Thus, in analogy with (3.23)

$$\alpha_T = \lim_{n \rightarrow \infty} \left(\frac{f_T^{(2^n)}(0)}{f_T^{(2^{n+1})}(0)} \right) = \lim_{n \rightarrow \infty} \left(\frac{f_T^{(2^n)}(0)}{f_T^{(2^{n+1})}(0)} \right)^2 = \alpha_T^2. \quad (3.42)$$

Similarly,

$$f_T^2 = \lim_{n \rightarrow \infty} (f_T^n), \quad (3.43)$$

and substitution of (3.41) leads to (3.38).

C. Tricritically: Crossover and scaling functions

The tricritical fixed-point discussion of the preceding section sets the stage for an examination of the structure of n cycles, etc. (the phase diagram) near the points T , T' , and their duals. It is a property of the iteration (3.3) that, if f_n has an m cycle, x_1, x_2, \dots, x_m , with even m , then f_{n+1} has two $m/2$ cycles, $x_1/f_n(1), x_2/f_n(1), \dots$, and $x_m/f_n(1), x_{m-1}/f_n(1), \dots$. If the m cycle of f_n is superstable, then both the $m/2$ cycles of f_{n+1} are superstable (f_{n+1} is at least bimodal). If m is odd, the f_{n+1} has an m cycle, $x_1/f_n(1), x_2/f_n(1), \dots, x_m/f_n(1), x_m/f_n(1), \dots$. Again, f_n superstable implies f_{n+1} superstable.

In the linear region (3.6)–(3.9) near a fixed point, these considerations take a particularly concrete form. Consider, for example, a function f_n with a 2^n cycle which lies near f_T^2 but not on the tricritical hypersurface. Such a function has the form

$$f_n = f_T^2 + \delta f,$$

with

$$\delta f(x) = \alpha_1 e_1(x) + \alpha_2 e_2(x), \quad (3.44)$$

where $e_1(x)$ and $e_2(x)$ are the eigenvectors belonging to $\delta^{(1)}$ and $\delta^{(2)}$, respectively. The iterate function (which has two 2^{n-1} cycles) is

$$f_{n+1} = f_T^2 + \delta f', \quad (3.45)$$

with

$$\delta f'(x) = \delta^{(1)} \alpha_1 e_1(x) + \delta^{(2)} \alpha_2 e_2(x),$$

i.e., it has moved away from the tricritical hypersurface by a factor $\delta^{(1)}$ in the even subspace and $\delta^{(2)}$ in the odd subspace [see (3.28) and (3.29)].

Now, consider the neighborhood of the tricritical point T . Any point in this neighborhood lies near the tricritical hypersurface and therefore maps after 2 iterations to the point T . The scaling behavior of f_T^2 is different from the scaling behavior of f_T because the scaling structure of f_T and f_T^2 is only in variables appropriately chosen by the in-

tervening iterations. These scaling variables are a natural set of coordinates near the point T . They may be taken to be

$$u_1 = b - b_T - a \tan \theta, \quad u_2 = a / \cos \theta, \quad (3.46)$$

where the slope of the critical line is

$$\tan \theta = -1.2347, \quad (3.47)$$

as illustrated in Fig. 16. Cycle structure near T has much regularity, when described in terms of u_1 and u_2 . For example, if there is a superstable 2^n cycle at (u_1, u_2) , then there is a superstable 2^{n+1} cycle at $(\delta_T^{(1)} u_1, \delta_T^{(2)} u_2)$, a superstable 2^{n+2} cycle at $(\delta_T^{(1)^2} u_1, \delta_T^{(2)^2} u_2)$, etc. provided only that all coordinates remain small enough so the points have not departed significantly from T (i.e., from the tricritical hypersurface).

These considerations allow a scaling²² description of the crossover from critical behavior along the Feigenbaum line to tricritical behavior along the b axis (region 2 of the scaling variables in Fig. 16). The position $u_1^{(n)}(u_2)$ of the superstable 2^n cycles in this region has a simple scaling form,

$$(\delta_T^{(1)})^n u_1^{(n)} = g((\delta_T^{(2)})^n u_2), \quad (3.48)$$

which is valid with a single function g independent of n , provided $n \gg 1$ and $|u_1|, |u_2| \ll 1$ but for arbitrary values of the scaled combinations $(\delta_T^{(2)})^n u_2$. Equation (3.48) must, of course, incorporate the known behavior on the b axis and along the critical line. On the b axis $u_2 = 0$, $u_1 = b - b_T$, so (3.19) gives

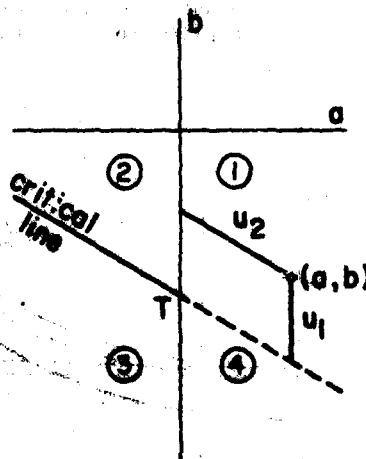


FIG. 16. Scaling variables for the tricritical point T . Equation (3.46) defines scaling variables (u_1, u_2) with reference to the point T , the b axis, and the critical line. In region 2 there is crossover between critical behavior along the critical line and tricritical behavior near the b axis.

$$\mathcal{F}(0) = B. \quad (3.49)$$

On the other hand, at fixed $u_2 \neq 0$ (3.48) must agree with (3.1) as $u_1 \rightarrow 0$, so at large x $\mathcal{F}(x) \sim x^\gamma$ and, comparing

$$u_1^{(n)} \sim (\delta_T^{(n)})^{-1} [(\delta_T^{(n)})^2 u_2]^\gamma \sim C/\delta_T^\gamma, \quad (3.50)$$

we deduce that the critical amplitude $C(a)$ vanishes as $a \rightarrow 0$ as a^γ , with

$$\gamma = \frac{\ln \delta_T^{(1)} - \ln \delta_T}{\ln \delta_T^{(2)}} = 0.4237, \quad (3.51)$$

which we have verified numerically.

Such scaling considerations are not restricted to region 3 (Fig. 16). Indeed, the n -cycle structure in the entire neighborhood of T scales in accordance with the discussion after (3.47). This point is illustrated by Figs. 17-20, which show the structure of 2^n cycles for $n=2, 3$, and 4 displayed in terms of the scaling combinations $(\delta_T^{(1)})^2 u_1$ and $(\delta_T^{(2)})^2 u_2$. The structure displayed should become fixed as $n \rightarrow \infty$. For $n=2$ some nonscaling features are observable far from the origin; however, $n=3$ and $n=4$ are already practically indistinguishable. Note that at fixed $\delta_T^2 u_2$, large k probes small u_1 ; thus, Figs. 17-20 illustrate graphically the high-order 2^n -cycle structures which pinch off the critical line at T . A similar discussion of behavior near T' , \tilde{T} , and \tilde{T}' could be made.

D. Other tricritical points

In this paper we have confined our discussion to the main Feigenbaum critical line TT' and its dual. We note in closing that tricritical structures exist at many other points in the a, b plane.

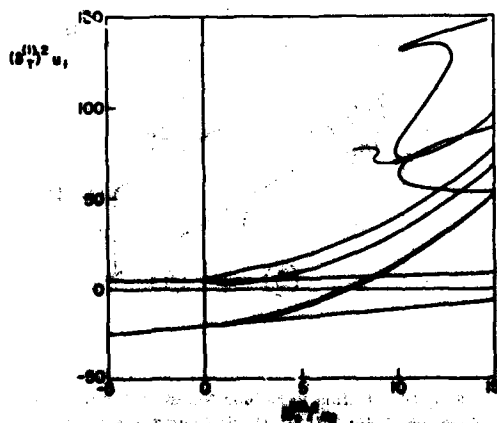


FIG. 17. Superstable 4-cycles near T plotted with scaling variables $(\delta_T^{(1)})^2 u_1$. Refer to the caption on Fig. 16. The point marked by the arrow is doubly stable; other scaling behavior is observable at the upper right.

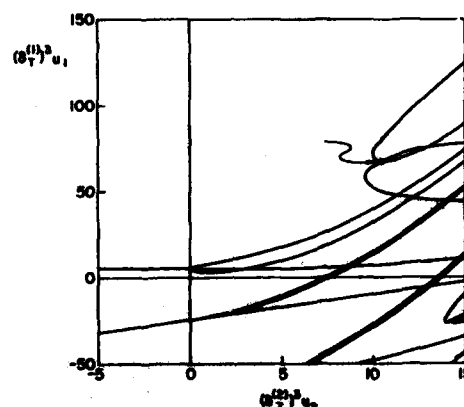


FIG. 18. Superstable 8-cycles near T plotted with scaling variables $(\delta_T^{(1)})^3 u_1$. The point marked by the arrow is doubly stable. This figure and Fig. 19 coincide except for small numerical differences.

Let us focus on the characterization (Sec. III A) of a tricritical point as the $n \rightarrow \infty$ limit of a set of doubly stable 2^n -cycle points. Consider only quadrant IV. The 2 cycles have a single doubly stable point (Fig. 8) at $(0, -1)$. The 4 cycles have (Fig. 11) two such points, $(0, -1.50393)$ and $(0.81176, -1.23953)$. The 8 cycles have four such points, etc. This sequence is illustrated in Fig. 21 through order $n=3$. It appears that, in going from n to $n+1$, each doubly stable point splits, forming as $n \rightarrow \infty$ a whole cascade of tricritical points. The outer sequence leads to T and \tilde{T}' . Others lie between T and \tilde{T}' . Each limiting point has an infinite number of others in its neighbor-

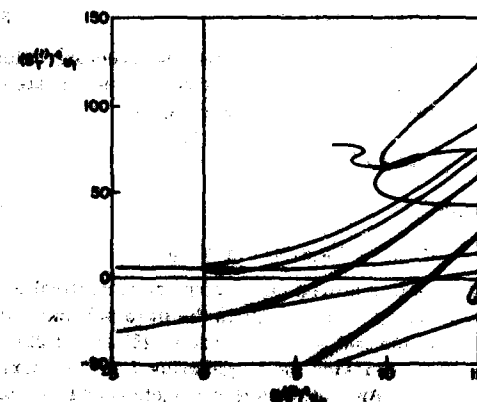


FIG. 19. Superstable 16-cycles near T plotted with scaling variables $(\delta_T^{(1)})^4 u_1$. The point marked by the arrow is doubly stable. This figure is identical, at numerical approximation to the extent $n \rightarrow \infty$ scaling function in this region of variation.

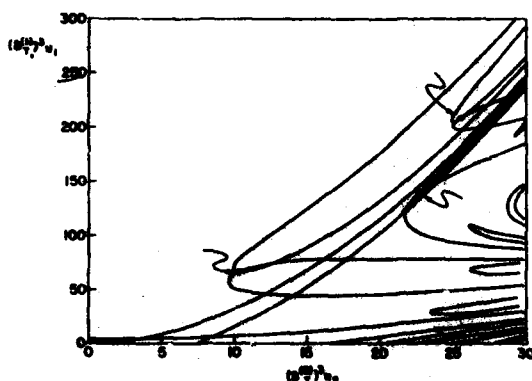


FIG. 20. Superstable 8 cycles associated with the extrema x_1 and x_2 (x_2 has been omitted for simplicity). Otherwise, this is the same as Fig. 18 except for the larger scale. Note the additional doubly stable points.

hood. Figures 17-19 show two sets of doubly stable 2^n -cycle points converging to T , one on the $x_2 = 0$ (b) axis and the other located at $(\delta_7^{(12)} x_1, \delta_7^{(12)} x_2) = (66.4, 10.3)$. Figure 20, which is on a larger scale, shows two additional doubly stable points, and it seems likely that there are many more at larger values of the scaling variables. The self-similarity observable in the scaling plots suggest that the tricritical points may form a Cantor set. At each level of n , we may label a 2^n doubly stable cycle mentioned above as right (R) or left (L) relative to its parent. We can then specify a tricritical point by a sequence of R 's and L 's. Conversely, any such a R - L sequence specifies a tricritical point. For instance, the tricritical point T is $R^n = (R, R, \dots)$ and the tricritical point T' is L^n . In Fig. 21, we have also plotted the locations of two additional tricritical points $T_2 = (0.35031, -1.7980)$ and $T_3 = (0.9704, -1.3767)$. In terms of the R - L notation, T_2 is RL^n and T_3 is

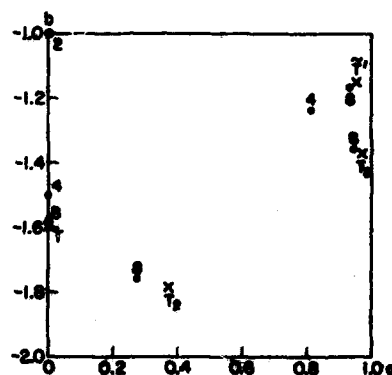


FIG. 21. Doubly stable 2^n cycles in the region between T and T' . The points T and T' are marked by crosses. In this region there are one 2 cycle, two 4 cycles, four 8 cycles, etc. All the $x \rightarrow \infty$ limit points are presumably tricritical.

LR^n . We have not at this stage located the critical lines associated with these tricritical points, nor have we explored the question of what fixed point they flow to under the iteration (3.3). Preliminary evidence shows other doubly stable points for the 8 cycles and above, which may well belong to additional tricritical sequences.

ACKNOWLEDGMENTS

We acknowledge with thanks the useful advice from and conversations with A. E. Jackson, L. P. Kadanoff, P. C. Martin, Y. Oono, and R. L. Schult. We thank L. H. Smith for help with computing. This work was supported in part by the National Science Foundation under Grants Nos. NSF DMR78-21066, NSF PHY79-00272, and by the Office of Naval Research under Contract No. N00014-80-C-0640.

¹R. M. May, *Nature* **261**, 459 (1976).

²P. Collet and J.-P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser, Basel, 1980); both Refs. 1 and 2 contain extensive references.

³M. Feigenbaum, *J. Stat. Phys.* **19**, 25 (1979); *Ibid.* **21**, 669 (1979).

⁴*Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 6.

⁵B. Derrida, A. Gervois, and Y. Pomeau, *J. Phys. A* **12**, 269 (1979).

⁶B. Derrida and Y. Pomeau, *Phys. Lett.* **224**, 317 (1980).

⁷Proceedings of CMLB Conference—Statistical Problems: Present and Future (North-Holland, in press).

⁸The relation is $a = \lambda(2 - \lambda)/4$, so $\lambda = 1$ (where a fixed point away from $x = 0$ first appears in the logistic map) is equivalent to $a = 1/4$, while $\lambda = 4$ (beyond which all of the unit intervals except a set of measure zero maps to $|x| = \infty$) is equivalent to $a = -2$.

⁹For a study of chaotic transitions and distributions among bands, see R. J. Chang and J. Wright, *Phys. Rev. A* **25**, 1419 (1981).

¹⁰The 2-unimodal maps of Ref. 2.

¹¹M. Misiurewicz, M. L. Sina, and P. R. Sina, *J. Combinatorial Theor.* **15**, 25 (1978).

¹²P. Collet, J.-P. Eckmann, O. E. Lanford III, *Commun. Math. Phys.* **75**, 311 (1980).

¹³Some of the results of this work are reported briefly in a short communication which will appear in Ref. 7.

S. J. Chang, M. Wortis, and J. Wright, Tricritical points and bifurcations in a quartic map (in press).

¹⁴For our purposes there is no need to distinguish between the ergodic and nonperiodic itineraries of Ref. 2.

¹⁵One-peak maps with this property are called "most stable" in Ref. 13. In this paper, we follow the convention of Refs. 2 and 5 and refer them as superstable. A cycle is called doubly stable if it is doubly superstable.

¹⁶The term "unimodal" is used in a similar but more restricted sense in Ref. 2.

¹⁷The converse is proved as follows: If the n cycle of f uses $x_0 = 0$, then there exists a y_0 such that $x_0 = G_2(G_2(y_0))$. Application of (2.14) to y_0 gives a \tilde{y}_0 which is a side extremum of f .

¹⁸A function $f(x)$ in the 1-cycle region may be unimodal or bimodal. As we shall see below, it is quite possible for a stable n cycle to coexist with a stable m cycle ($n \neq m$ or $n = m$) or with chaotic trajectories.

¹⁹There are many other Feigenbaum critical lines in the phase diagram. The "main" critical lines shown in Fig. 10 have a complete bifurcation sequence 1, 2, 4, 8, The other critical lines have truncated sequences 2, 4, 8, ..., 4, 8, 16, ..., and so forth.

²⁰Note, for example, that the lower of the two main diagonal 4-cycle lines in Fig. 11 crosses the negative a axis beyond a_0 . It corresponds to the "chaotic 4 cycle," which occurs in the region of inverse bifurcation.

²¹If the quadratic map (1.2) is a good guide, then we may expect that the n -cycle regions are dense (Ref. 2), even though the chaotic regions have finite measure (conjectured in Ref. 2; proven by M. Jakobson (unpublished)).

²²D. Singer, *SIAM (J. Appl. Math.)* **35**, 260 (1978).

²³When (2.25) is satisfied in quadrant III, the whole measure of the iterative domain of f is apparently exhausted by either a single stable n cycle or a single

region of chaotic itineraries. In principle, there are two additional possibilities: simultaneous existence of two or more distinct chaotic regions or coexistence of one stable n cycle with one or more distinct chaotic regions. These possibilities do not seem to occur in practice; however, to the best of our knowledge they have not been rigorously ruled out, even for the logistic map.

²⁴Again, there exists the logical possibility that f possesses some nonzero region of chaotic itineraries. So far as we know this does not occur; however, we are aware of no rigorous argument to exclude it.

²⁵In the theory of phase transitions a tricritical point marks the point where a line of second-order transitions (a critical line) becomes first order. It is characterized by the appearance of a new fixed point with an extra relevant variable. See R. B. Griffiths, *Phys. Rev. Lett.* **24**, 715 (1970); E. K. Riedel, *ibid.* **32**, 675 (1972); E. K. Riedel and F. J. Wegner, *ibid.* **29**, 349 (1972). Let the reader be warned, however, that the parallel is imperfect: What happens in the bifurcation problem beyond the tricritical point is not analogous to first-order behavior. Furthermore, there is not here (as there would be in the phase-transition problem) an enlarged parameter space containing three critical lines and a triple line meeting at tricriticality.

²⁶We follow here the phase-transition terminology of, for example, Ref. 4.

²⁷One method (Ref. 3) is to substitute (3.26) into the fixed-point equation (3.4) and then truncate to some given order. A simple two-coefficient truncation gives $c_1^{(2)} = -1.855$, $c_2^{(2)} = 0.278$, and $\delta_2^{(2)} = 0.23$, in reasonable agreement with the accurate results (3.19) and Table II.

²⁸An early example of this method was the analysis by J. T. Ho and J. D. Litster, *Phys. Rev. Lett.* **22**, 603 (1969), of the magnetic equation of state of CrBr_2 .

ILL-(TH)-82-20
June 1982

SPATIAL ORDER IN CHAOS

R.L. Schult and Jon Wright*

Department of Physics
University of Illinois at Urbana-Champaign
1110 West Green Street
Urbana, IL 61801

Present Address: * La Jolla Institute
P. O. Box 1434
La Jolla, CA 92038

PACS Index: 02.50.+s
47.25.-c
03.40.-t
05.40.+j

Typed by Mary Ostendorf

Abstract

Highly damped dynamical systems with low-dimensional strange attractors have temporal features which can be described by one-dimensional maps. We demonstrate that there should also be spatial features describable by one-dimensional maps and suggest experimentalists investigate certain spatial correlations.

APR 28 1980
 1-25-80
 1-25-80
 1-25-80

Revised 1-25-80

In this paper we discuss some of the phenomenology of order and chaos in highly damped dynamical systems, with emphasis on the kinds of information that can be obtained experimentally. Our main concern is with systems having a low dimensional strange attractor. The particular examples chosen correspond to an attractor with fractal dimension very close to two. As is well known from the work of Lorenz¹⁾, such systems have important temporal features that can be described by one dimensional maps. We wish to point out that there will also be spatial correlations describable by one dimensional maps. It does not seem to be widely appreciated how much information can be obtained from such maps, and one of the purposes of this paper is to illustrate the kinds of information that can be obtained and to suggest that certain experiments could be performed to test some of the ideas presented here.

We will use the Lorenz model¹⁾ to illustrate our points, but we have also done computer experiments on a model with 5 modes to verify that 3 modes was not special. Our main point will concern spatial correlations in chaotic behavior implied by the existence of low dimensional attractors. To demonstrate that it will be necessary first to discuss temporal correlations implied by the existence of one dimensional maps.

The Lorenz equations are given by:

$$\dot{X} = Y - \sigma X$$

$$\dot{Y} = -XZ + Y - \epsilon Y$$

$$\dot{Z} = XY - \epsilon Z$$

where we have rescaled the variables as discussed in ref. 2. The figures are

for $\epsilon = .02216$, $\sigma = 10.$, and $B = 8/3$ which correspond to $r = 1/(\epsilon^2 \sigma) = 203.64$ in Lorenz's¹⁾ notation. The particular parameters used correspond to the chaotic region following the first period doubling regime as ϵ is increased from zero; however very similar behavior is obtained for the parameters studied by Lorenz.

The time series for the variable X is shown in Figure 1. Pictures of the three dimensional orbits (see refs. 1, 3-4) show that the motion is nearly confined to a two dimensional surface — although it is infinite sheeted. If the attractor were in fact two dimensional we could predict everything about the system from a knowledge of any two variables which designate the position on this two-dimensional surface. One choice of these two variables could be X and \dot{X} . We construct our one dimensional maps by considering various quantities at the time when X has an extremum, (i.e. $\dot{X} = 0$), although any other reference point would suffice. (Some choices of reference points give simpler maps than others.) Examples are given in Figures 2 through 4. Figure 2 is similar to Lorenz's¹⁾ Figure 4 and shows the predictability of the value of X at its next maximum from the value $X = X_{MAX}$, at the present maximum. See Lorenz's¹⁾ Figure 3 for another example. Note that there is a nonuniqueness that we will discuss later. Figure 3(a) shows the predictability of the time to the next maximum of X from the present maximum. Figure 3(b) shows the predictability of the value of Z at the next maximum of X . Figure 3(c) shows the value of Z at the time of the present maximum of X . In all of the examples, the present value of X at its maximum $X = X_{MAX}$ serves as a predictor.

Thus we expect that if the data from a single probe experiment gives a one dimensional map such as that shown in figure 2, then a two probe experiment should provide a map such as that shown in figure 3(c) (or figure

3(b)). To date no experimentalists have looked at such maps and we suggest that they do so.

One might ask if it is possible to predict at times other than extrema and the answer is yes. Figure 4 shows the prediction of Z at a time $\Delta t = 20$ after the maximum of X . Other values of Δt give similar curves. It is of course not possible to take Δt arbitrarily large. There are two reasons for this, one of which is that these one dimensional maps display what is called sensitive dependence on initial conditions so that slight differences in X_{MAX} are greatly magnified at large Δt giving the map a complicated structure which is difficult to resolve. The second is that the map is really not one dimensional, but the lines all are fractal structures, or from the point of view of an experimentalist they are thick lines. This thickness introduces slight errors which are then magnified by the sensitive dependence on initial conditions.

We now discuss briefly the nonuniqueness. Typically there will be a multivalued problem that is easily resolved by a simple rule. In Figure 2 we have divided the curve into four regions separated by *. If we are in the region where we have to select either the curve C or D we do so by the following rule. Look at the previous point. If it is was on D or A use C. If it was on C or B use D. The rule is easily obtained by watching the generation of a sequence of a few points. We have looked at a number of systems and the appropriate rule is usually easy to construct.

Here we have ignored the thickness of the lines. If we resolve them, at the next level each segment becomes two and our rule becomes more complicated, but still tractable. This labelling can be continued to any depth desired, but in practice an experimentalist can only distinguish a few curves. Note that the previous value is not needed, only an integer label. This same label serves to label the other curves such as those shown in Figures 3-4.

There have been several attempts to discuss the statistics⁴⁻⁷⁾ of systems of equations such as the Lorenz equations. Although we have nothing new to add here, it should be pointed out that the maps we have shown imply that there are strong temporal correlations of some sort in such systems. Typically the autocorrelation function falls off much faster in time than the memory of the maps. Standard statistical descriptions don't correctly treat this feature. Some different type of description is needed which takes into account the existence of these maps.

Finally we turn to the main point of this paper — the possibility of spatial order in chaos. Suppose one has a system described by partial differential equations. If the system is expanded in terms of some complete set of modes, and the number of modes truncated, ordinary differential equations are obtained. Now if, as we saw above, all of the mode amplitudes are in fact given in terms of two modes (plus a label) at any time, then it follows that the original field variables are also given in terms of two mode amplitudes. In fact the value of one field, say temperature, at one point at a time when this field is an extremum should serve as a predictor for field variables at the same or different spatial points. To illustrate the idea let us return to the Lorenz equations. The temperature field $\theta(\vec{r}, t)$ is given in terms of Y and Z and the stream function $u(\vec{r}, t)$ in terms of X .

$$\theta(\vec{r}, t) = K_1 (\sqrt{2} Y(t) \sin \pi x_3 \cos kx_1 + Z(t) \sin 2\pi x_3)$$

$$u(\vec{r}, t) = K_2 X(t) \sin \pi x_3 \sin kx_1$$

where, K_1 , K_2 , and k are constants and $\vec{r} = (x_1, x_2, x_3)$.

In Figure 5 we show the use of the temperature at one spatial point at the time of a temperature extremum at that point as a predictor for the temperature at a different spatial point. The points were chosen (at random) so that the predictor is $\theta_1 = .134 Y(t) + .6Z(t)$ and the predicted temperature is $\theta_2 = .5Z(t) - .146Y(t)$. The temperature at its extremum at one spatial point also serves as a predictor for the velocity field at any other point.

We suggest that it should be possible to test these ideas by performing experiments — both real and on the computer. Of course, our examples have been chosen to show our ideas in the most favorable light. Real systems may have a more complicated sheet structure that is hard to untangle — but if one is convinced there is a rule and if the maps appear to be one dimensional it should be possible.

Specifically, choose any two quantities, $U(t)$ and $V(t)$. These could be, for example, the values of some field at two spatially distinct points at the same time. Plot $U(t)$ vs $V(t)$ for those values of time for which $\dot{U}(t) = 0$. If this plot shows lines as in Figures 2-5 the dimension of the attractor is near two. If the plot is scattered a higher dimension is indicated.

In systems with attractors near three dimensions it should be possible to make contour plots of the value of some variable at one point as a function of the values at two other points (at the time when one of them is an extremum). Note here we need the values of three quantities to predict others. The plots would present information in the form of Lorenz's¹⁾ Figure 3 except that there is the additional information that another variable has prescribed value (i.e. $\dot{U} = 0$).

Given a map such as figure 2, we can estimate something about how large A_t (see figure 4) can be before temporal information is lost. At present we do not know how to make the corresponding estimates for spatial structure.

Experimental information where one probe is held fixed and a second probe is moved to obtain data at several different spatial points would be useful.

At this point we would like to mention the possibility of using such maps for design purposes. By changing parameters slightly it is possible to see how the map is changing, and by using a graphical Newton's method experimentalists could predict the parameters that give a particular desired behavior, for example a periodic orbit. We have used this scheme successfully on the Lorenz equations to locate periodic orbits with particular symmetry properties.

We conclude that in systems with low dimensional attractors there should be highly correlated spatial structures. This could be demonstrated and investigated by making maps (or contour plots) similar to those already used for displaying temporal correlations.

Supported in part by ONR, N00014-80-C-0840.

References

1. E. N. Lorenz, J. Atmos. Sci. 20, 130 (1963).
2. K. A. Robbins, SIAM J. Appl. Math. 36, 457 (1979).
3. T. Shimizu and N. Morioka, Phys. Lett. 66A, 182 (1978), 66A, 447 (1978).
4. R. Graham and H. J. Scholz, Phys. Rev. A22, 1198 (1980).
5. S. Grossman and B. Sonneborn-Schmick, Phys. Rev. 25, 2371 (1982).
6. M. Lucke, J. Stat. Phys. 15, 455 (1976).
7. A. Zippelius and M. Lucke, J. Stat. Phys. 24, 345 (1981).

Figure Captions

- 1 Segment of the time series for the variable X for the Lorenz equations for the parameters $c = .02216$, $\sigma = 10.$, $\beta = 8/3$.
- 2 Map showing the value of X at the next maximum of X (MX) given the value of X at the present maximum of X . Segments A,B,C,D are discussed in the text.
3. Maps showing: a) the time to the next maximum of X (MX) b) the value of Z at MX and c) the value of Z at the present maximum of X (PMX) as a function of X at PMX .
4. Map showing Z at a time $\Delta t = 20$ after PMX as a function of X at PMX .
5. Map showing the temperature, θ_2 , at one spatial point as a function of the temperature, θ_1 , at another point, given that θ_1 is maximal.

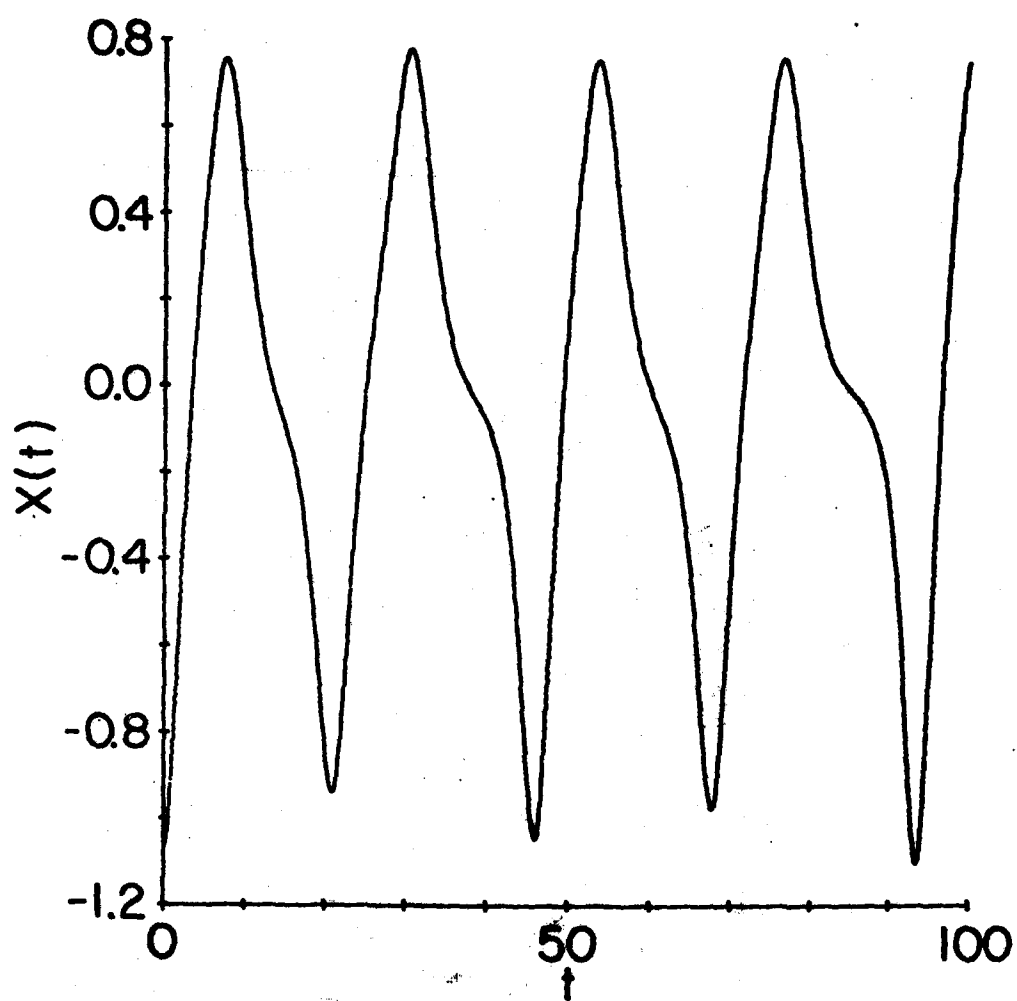


Fig. 1

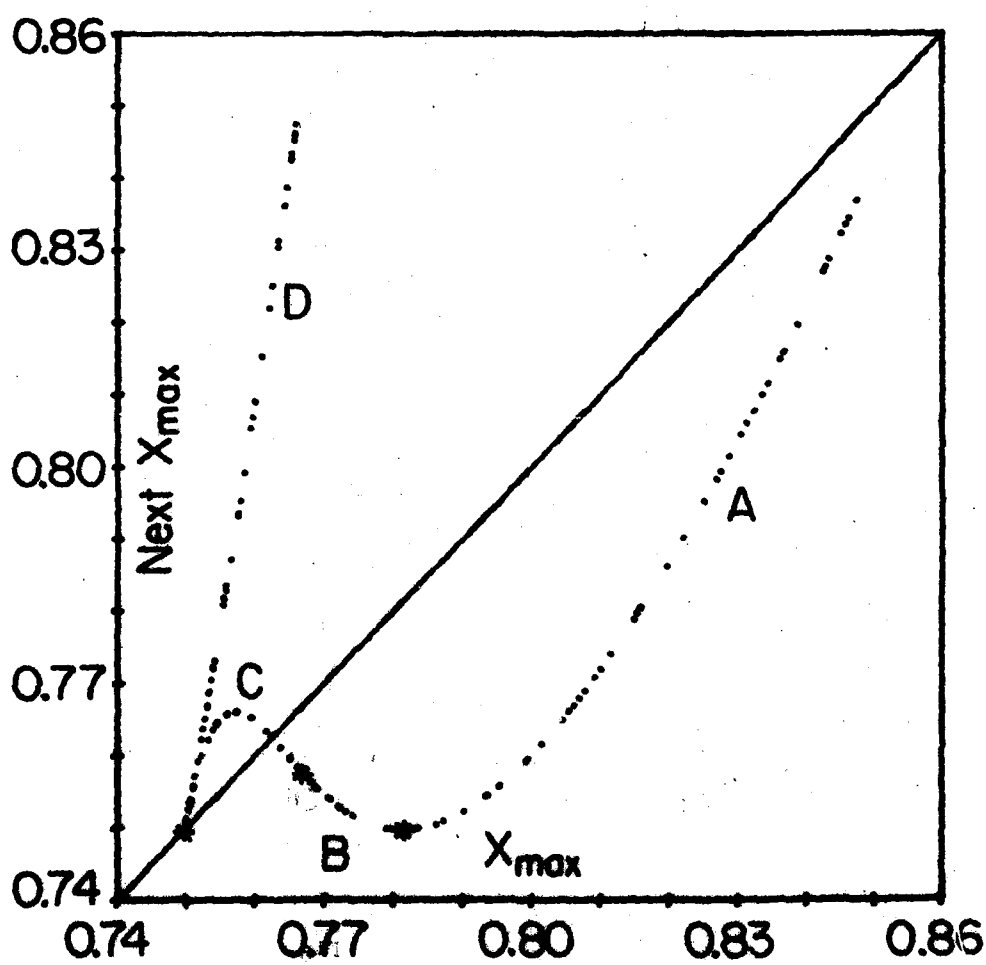


Fig. 2

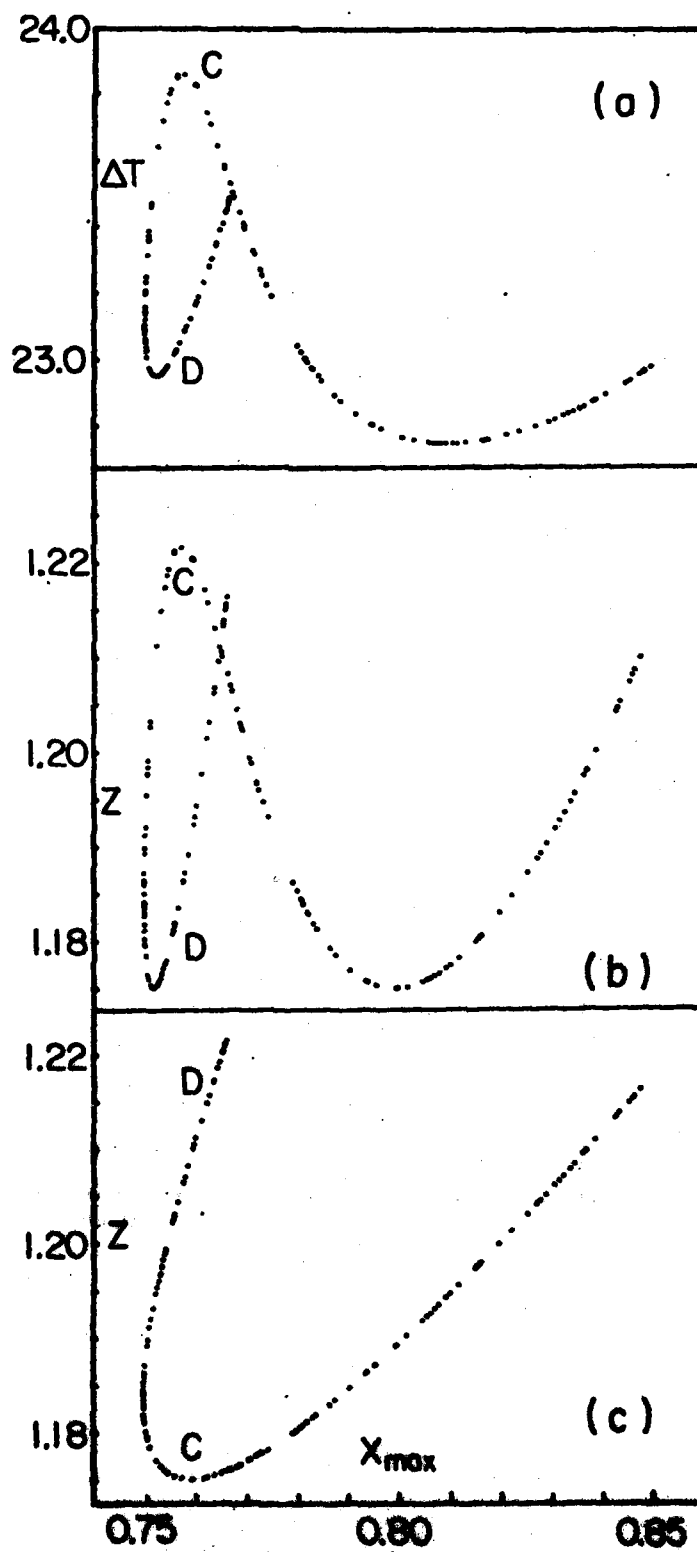


Fig. 3a,b,c

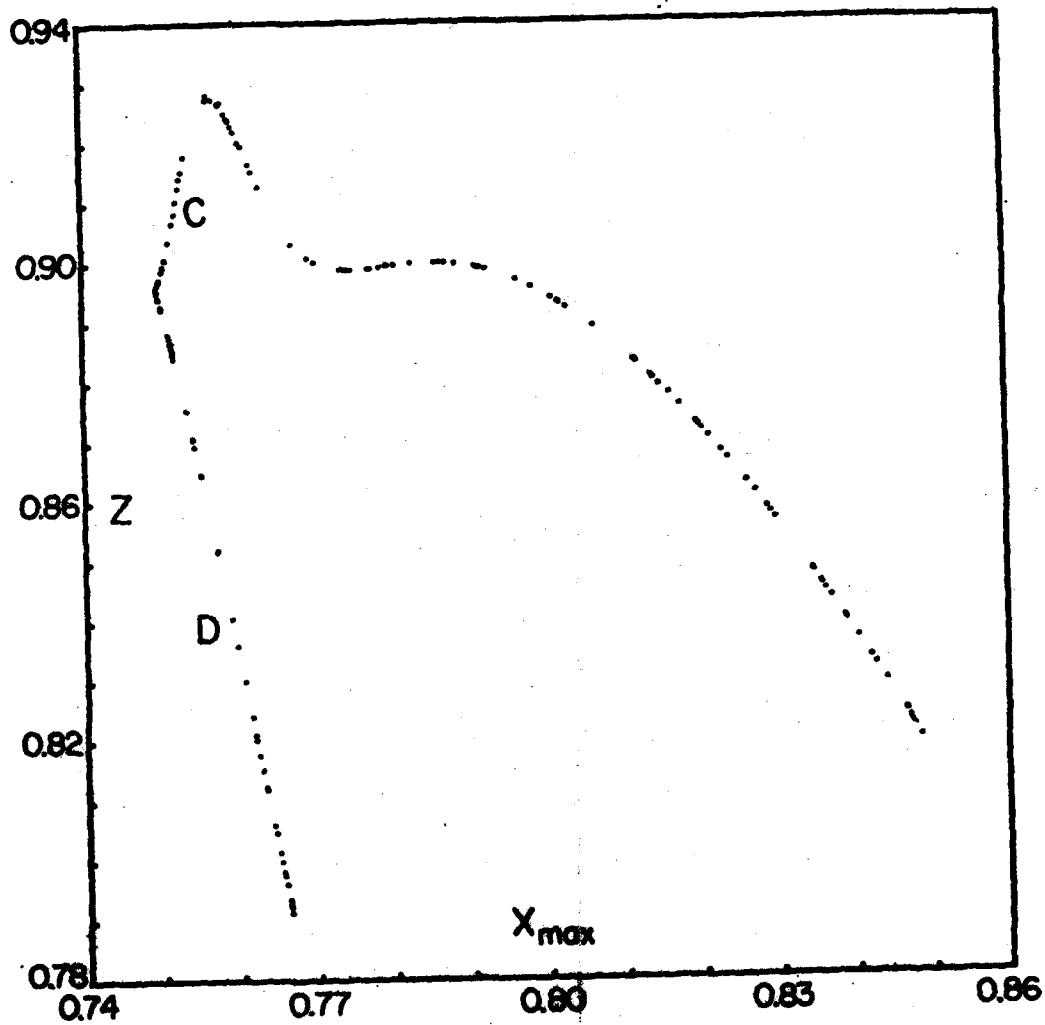


Fig. 4

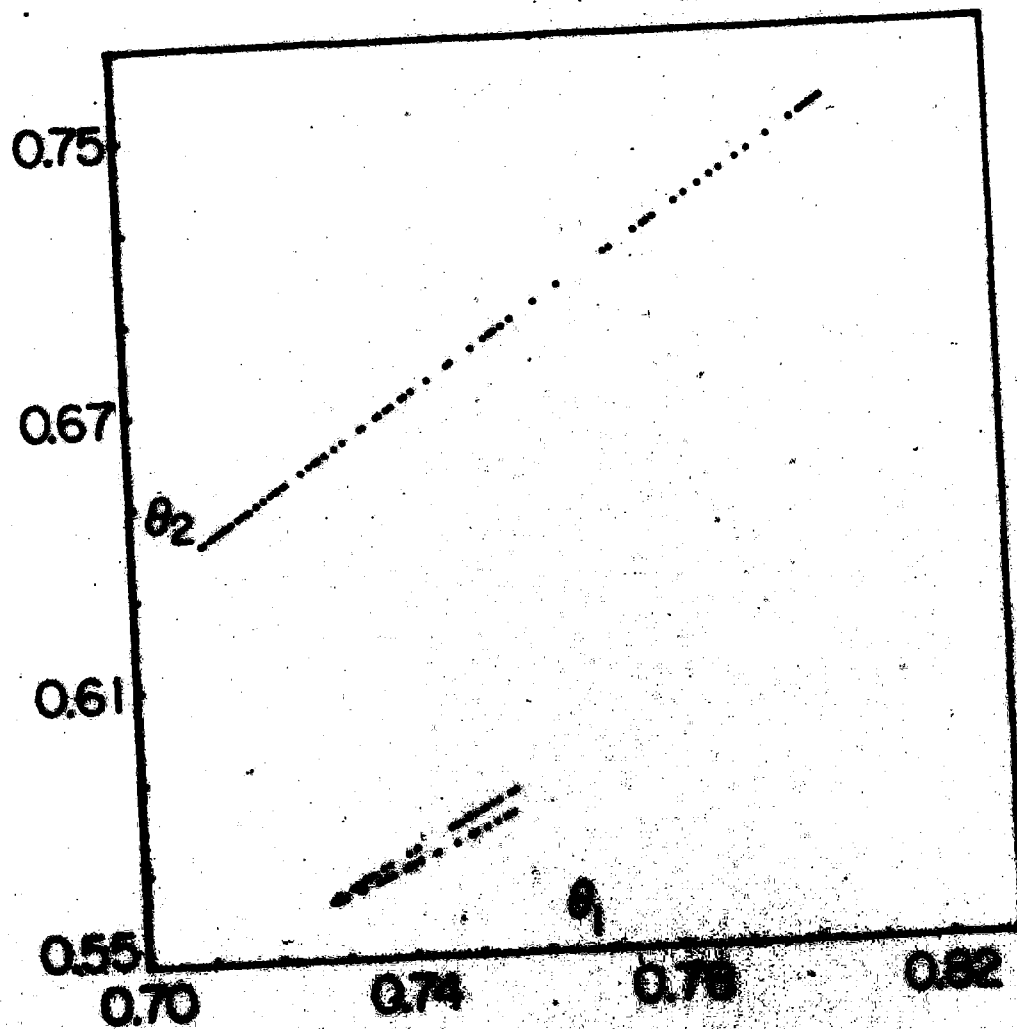


Fig. 3